Vol. 42, No. 3, July, 1941
ANNALS OF MATHEMATICS

ON POISSON'S SUMMATION FORMULA

BY A. P. GUINAND

(Received October 6, 1939)

1. INTRODUCTION

A number of papers have been written in recent years on various examples and applications of the formula

$$\frac{1}{2}f(0) + \sum_{n=1}^{\infty} f(n) = \int_0^{\infty} f(t)\,dt + 2\sum_{n=1}^{\infty}\int_0^{\infty} \cos 2\pi ntf(t)\,dt. \tag{1.1}$$

This formula is usually known as Poisson's summation formula. Other forms of the formula are

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} \cos 2\pi ntf(t)\,dt, \tag{1.2}$$

and

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} e^{2\pi int}f(t)\,dt. \tag{1.3}$$

Further, if we write

$$g(x) = 2\int_0^{\infty} \cos 2\pi xtf(t)\,dt, \tag{1.4}$$

then, with suitable conditions,

$$f(x) = 2\int_0^{\infty} \cos 2\pi xtg(t)\,dt \tag{1.5}$$

by Fourier's integral theorem, and (1.1) can be written

$$\frac{1}{2}f(0) + \sum_{n=1}^{\infty} f(n) = \frac{1}{2}g(0) + \sum_{n=1}^{\infty} g(n). \tag{1.6}$$

Several writers have pointed out that it is often difficult to justify the application of the formula even in cases where the formal result can be shown by other methods to be true.[1] Most methods which have been used for proving such formulae are elaborations of one of the two following types of formal procedure.

[1] Erdelyi (1), 408, Kober (1), 612.

(A). We may regard (1.2) as a Fourier series.[2] If we expand the periodic function

$$\sum_{n=-\infty}^{\infty} f(n+x)$$

as a Fourier series for $-\frac{1}{2} < x < \frac{1}{2}$, and then put $x = 0$ we formally obtain (1.2).
(B). We may regard (1.3) as a result in the theory of analytic functions.[3] If $f(z)$ is analytic, and we integrate $\pi \cot \pi z f(z)$ around the rectangle R with vertices $N + \frac{1}{2} \pm ai$, $-N - \frac{1}{2} \pm ai$, then by Cauchy's theorem

$$\frac{1}{2\pi i} \int_R \pi \cot \pi z f(z)\, dz = \sum_{n=-N}^{N} f(n).$$

Putting

$$\pi \cot \pi z = \pi i + 2\pi i \sum_{n=1}^{\infty} e^{-2n\pi i z}$$

when

$$I(z) < 0,$$

and

$$\pi \cot \pi z = -\pi i - 2\pi i \sum_{n=1}^{\infty} e^{2n\pi i z}$$

when

$$I(z) > 0,$$

and integrating term by term, we formally deduce (1.3) on making N tend to infinity and a to zero.

In addition to these methods two further methods have recently been developed which admit of application to functions having various types of singularities.
(C). Ferrar[4] derives a generalization of (1.1) by the use of Mellin transforms and the functional equation of the Riemann zeta-function. Conditions are applied to the Mellin transform of $f(x)$ rather than to $f(x)$ itself.
(D). In a recent paper[5] I gave another method for application to more general summation formulae of which (1.1) is a particular case. In this method we regard (1.1) as a Parseval theorem in the theory of Hankel transforms of functions of the Lebesgue class $L^2(0, \infty)$, and prove:

If $f(x)$ is an integral, and both $f(x)$ and $xf'(x)$ belong to $L^2(0, \infty)$, then

$$\lim_{N\to\infty} \left\{ \sum_{n=1}^{N} f(n) - \int_0^N f(t)\, dt \right\} = \lim_{N\to\infty} \left\{ \sum_{n=1}^{N} g(n) - \int_0^N g(t)\, dt \right\}, \tag{1.7}$$

[2] Courant and Hilbert (1), 64–65.
[3] Kronecker (1), Landsberg (1).
[4] Ferrar (1).
[5] Guinand (1). Put $a_n = 1$ in Theorem 2.

where $g(x)$ satisfies

$$\int_0^x g(y)\,dy = \frac{1}{\pi}\int_0^\infty \frac{\sin 2\pi xy}{y} f(y)\,dy$$

and is chosen so that it is the integral of its derivative.

The formula (1.7) obviously agrees with (1.6) if the series and integrals concerned converge. However, this is not necessary for the truth of (1.7), and it covers a number of examples in which (1.6) is inapplicable.

In the present paper we simplify the above result slightly, and extend it to functions of the Lebesgue classes $L^p(0, \infty)$, $(1 < p \leqq 2)$. We also give examples to show the ease with which the result can be applied, and discuss extensions to sums involving primitive characters.

2. Preliminary Lemmas

Lemma 1.[6] *If $f(x)$ belongs to $L^p(0, \infty)$ for some p satisfying $1 < p \leqq 2$, and*

$$p' = p/(p-1), \qquad R(\nu) > -1,$$

then there exists a function $g(x)$ belonging to $L^{p'}(0, \infty)$ defined almost everywhere by

$$g(x) = \operatorname*{l.i.m.}_{T\to\infty} 2\pi \int_0^T (xt)^{\frac{1}{2}} J_\nu(2\pi xt) f(t)\,dt. \tag{2.1}$$

Further

$$f(x) = \operatorname*{l.i.m.}_{T\to\infty} 2\pi \int_0^T (xt)^{\frac{1}{2}} J_\nu(2\pi xt) g(t)\,dt, \tag{2.2}$$

and if $f_1(x)$, $g_1(x)$ and $f_2(x)$, $g_2(x)$ are two pairs of such transforms for the same values of p and ν, then

$$\int_0^\infty f_1(x)g_2(x)\,dx = \int_0^\infty f_2(x)g_1(x)\,dx. \tag{2.3}$$

Lemma 2. *If $f(x)$ is an integral and tends to zero as x tends to infinity, and $xf'(x)$ belongs to $L^p(0, \infty)$, $(p > 1)$ then $f(x)$ belongs to $L^p(0, \infty)$, and $x^{1/p}f(x)$ tends to zero as x tends to $+0$ or infinity.*

Hardy[7] has proved that, if $\psi(x)$ belongs to $L^p(0, \infty)$, so does

$$\varphi(x) = \int_x^\infty \psi(t)\,\frac{dt}{t},$$

and

$$x^{1/p}\varphi(x) \to 0 \text{ as } x \to +0 \text{ or } x \to \infty.$$

Putting

$$\psi(x) = xf'(x)$$

the result follows immediately.

---

[6] Busbridge (1), or Kober (2). We only require the result for $\nu = \frac{1}{2}$.

[7] Titchmarsh (1), 396.

LEMMA 3.[8] *The function* $x^{-1}\{[x] - x\}$ *is self-reciprocal with respect to the kernel* $2\pi x^{\frac{1}{2}}J_{\frac{1}{2}}(2\pi x)$. *That is*

$$2\pi \int_0^{\infty} \{[t] - t\}\, x^{\frac{1}{2}} t^{-\frac{1}{2}} J_{\frac{1}{2}}(2\pi xt)\, dt = x^{-1}\{[x] - x\}.$$

LEMMA 4. *If* $f(x)$ *satisfies the conditions of Lemma* 2 *then it has a transform* $g(x)$ *with respect to the kernel* $2\cos 2\pi x$ *given for* $x > 0$ *by*

$$g(x) = 2\int_0^{\infty} \cos 2\pi xt f(t)\, dt. \tag{2.4}$$

Further

$$f(x) = 2\int_0^{\infty} \cos 2\pi xt g(t)\, dt,$$

$g(x)$ *is an integral, both* $g(x)$ *and* $xg'(x)$ *belong to* $L^{p'}(0, \infty)$, $x^{1/p'}g(x)$ *tends to zero as* x *tends to* $+0$ *or* ∞, *and* $xf'(x)$, $xg'(x)$ *are a pair of transforms with respect to the kernel* $2\pi x^{\frac{1}{2}}J_{\frac{1}{2}}(2\pi x)$ *in the sense of Lemma* 1.

Put

$$G_1(x) = \underset{T\to\infty}{\text{l.i.m.}}\ 2\int_0^T \cos 2\pi xt f(t)\, dt$$

and

$$G_2(x) = \underset{T\to\infty}{\text{l.i.m.}}\ 2\pi \int_0^T (xt)^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi xt) t f'(t)\, dt.$$

By Lemmas 1 and 2 it follows that these expressions exist almost everywhere, and $G_1(x)$, $G_2(x)$ both belong to $L^{p'}(0, \infty)$. Now, if $x > 0$

$$\begin{aligned}\int_x^{\infty} |\, G_2(u)\, |\, \frac{du}{u} &\leqq \left\{\int_x^{\infty} |\, G_2(u)\, |^{p'}\, du\right\}^{1/p'} \left\{\int_x^{\infty} u^{-p}\, du\right\}^{1/p} \\ &\leqq \left\{\int_0^{\infty} |\, G_2(u)\, |^{p'}\, du\right\}^{1/p'} \left\{\frac{x^{1-p}}{p-1}\right\}^{1/p}.\end{aligned}$$

Hence the integral converges absolutely, and

$$\begin{aligned}\int_x^{\infty} G_2(u)\, \frac{du}{u} &= 2\pi \int_x^{\infty} \frac{du}{u}\, \underset{T\to\infty}{\text{l.i.m.}} \int_0^T (ut)^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi ut) t f'(t)\, dt \\ &= 2\pi \lim_{T\to\infty} \int_x^{\infty} u^{-\frac{1}{2}}\, du \int_0^T t^{\frac{1}{2}} J_{\frac{1}{2}}(2\pi ut) f'(t)\, dt \\ &= 2\pi \lim_{T\to\infty} \int_0^T t^{\frac{1}{2}} f'(t)\, dt \int_x^{\infty} u^{-\frac{1}{2}} J_{\frac{1}{2}}(2\pi ut)\, du\end{aligned}$$

[8] Titchmarsh (2), 267. We also use the notation

$$\int_a^{\infty} = \lim_{T\to\infty} \int_a^T.$$

$$= \frac{1}{\pi x} \lim_{T\to\infty} \int_0^T \sin 2\pi x t f'(t)\, dt$$

$$= \frac{1}{\pi x} \lim_{T\to\infty} \left\{ [\sin 2\pi x t f(t)]_0^T - 2\pi x \int_0^T \cos 2\pi x t f(t)\, dt \right\}$$

$$= -2 \int_0^{\to\infty} \cos 2\pi x t f(t)\, dt.$$

The inversion of order of integration is justified by absolute convergence, and the integrated terms vanish by Lemma 2. Hence the limit (2.4) exists, and we may put

$$g(x) = -\int_x^\infty G_2(u) \frac{du}{u} = 2 \int_0^{\to\infty} \cos 2\pi x t f(t)\, dt$$

for all positive x. Hence $g(x)$ is an integral and $G_2(x) = xg'(x)$ almost everywhere. I.e. $xf'(x)$, $xg'(x)$ are transforms of the required type, and the rest of the lemma follows from Lemma 2.

3. The Summation Formula

We can now prove:

THEOREM 1. *If $f(x)$ is an integral, tends to zero at infinity, and $xf'(x)$ belongs to $L^p(0, \infty)$, $(1 < p \leqq 2)$ then*

$$\lim_{N\to\infty} \left\{ \sum_{n=1}^N f(n) - \int_0^N f(t)\, dt \right\} = \lim_{N\to\infty} \left\{ \sum_{n=1}^N g(n) - \int_0^N g(t)\, dt \right\} \tag{3.1}$$

where

$$g(x) = 2 \int_0^{\to\infty} \cos 2\pi x f(t)\, dt.$$

Now $x^{-1}\{[x] - x\}$ belongs to $L^p(0, \infty)$ for any $p > 1$, so by Lemmas 3 and 4 and (2.3)

$$\int_0^\infty x^{-1}\{[x] - x\}\, xf'(x)\, dx = \int_0^\infty x^{-1}\{[x] - x\}\, xf'(x)\, dx.$$

The left-hand side is

$$\int_0^\infty \{[x] - x\}\, f'(x)\, dx = (\{[x] - x\}\, f(x)]_0^\infty - \int_0^\infty f(x)\, d\{[x] - x\}$$

$$= -\lim_{N\to\infty} \int_0^N f(x)\, d\{[x] - x\}$$

$$= -\lim_{N\to\infty} \left\{ \sum_{n=1}^N f(n) - \int_0^N f(t)\, dt \right\}.$$

Treating the right-hand side in the same way, we have (3.1).

However, Theorem 1 does not cover functions having discontinuities, for $f(x)$, being an integral, is necessarily continuous. In order to obtain a sufficiently general theorem covering such functions we combine Theorem 1 with a known form of Poisson's formula.

THEOREM 2. *If $f(x)$ is of bounded variation in $(0, \infty)$, and tends to zero at infinity, then*

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N}\tfrac{1}{2}\{f(n+0)+f(n-0)\}-\int_0^N f(t)\,dt\right]=\sum_{n=1}^{\infty}g(n)-\int_{-0}^{\infty}g(t)\,dt,$$

where

$$g(x)=2\int_{-0}^{\infty}\cos 2\pi xt\,f(t)\,dt$$

exists as a Cauchy integral at zero and infinity.

Titchmarsh[9] has shown that with these conditions

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N}\tfrac{1}{2}\{f(n+0)+f(n-0)\}-\int_0^{N+\frac{1}{2}} f(t)\,dt\right]=\sum_{n=1}^{\infty}g(n)-\tfrac{1}{2}f(0+0).$$

Obviously

$$\lim_{N\to\infty}\int_N^{N+\frac{1}{2}} f(t)\,dt=0,$$

and we can show that[10]

$$\tfrac{1}{2}f(0+0)=\int_{-0}^{\infty}g(t)\,dt.$$

Theorem 2 follows immediately.

Now it sometimes happens that a function $f(x)$ satisfies the conditions of Theorem 1 for one set of values of p in the neighborhood of the origin and for another set of values at infinity. In order to cover all these possibilities we will take as our general theorem:

THEOREM 3. *If $f(x)$ can be expressed as the sum of three functions $f_1(x)$, $f_2(x)$, $f_3(x)$ such that all three tend to zero at infinity, $f_1(x)$ and $f_2(x)$ are integrals, $xf_1'(x)$ belongs to $L^{p_1}(0,\infty)$ for some p_1 in $(1<p_1\leqq 2)$, $xf_2'(x)$ belongs to $L^{p_2}(0,\infty)$ for some p_2 in $(1<p_2\leqq 2)$, and $f_3(x)$ is of bounded variation in $(0,\infty)$, then*

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N}\tfrac{1}{2}\{f(n+0)+f(n-0)\}-\int_0^N f(t)\,dt\right]$$

$$=\lim_{N\to\infty}\left[\sum_{n=1}^{N}g(n)-\int_{-0}^{N}g(t)\,dt\right], \tag{3.2}$$

[9] Titchmarsh (2), Theorem 45.

[10] Titchmarsh (2), Theorem 6. The required extension to the case $x=0$ is trivial.

where

$$g(x) = 2\int_{-0}^{+\infty} \cos 2\pi xt\, f(t)\, dt.$$

(3.2) holds for $f_1(x)$, $f_2(x)$ and $f_3(x)$ separately, and Theorem 3 follows by addition.

4. Examples

Many examples of Poisson's summation formula have been noted by various writers.[11] Most of these examples can be justified by Theorem 3 without difficulty. The examples given below are chosen to show the advantage of the use of Poisson's formula in the form (1.7) rather than in any of the forms (1.1), (1.2), or (1.3).

(i). Put

$$f(x) = x^{-s} \qquad (\tfrac{1}{2} < s < 1).$$

The conditions of Theorem 3 are satisfied if we put

$$f(x) = f_1(x) + f_2(x),$$

$$f_1(x) = \begin{cases} x^{-s} & (x < 1) \\ e^{1-x} & (x \geqq 1), \end{cases}$$

$$f_2(x) = \begin{cases} 0 & (x < 1) \\ x^{-s} - e^{1-x} & (x \geqq 1), \end{cases}$$

$$1 < p_1 < s^{-1}, \qquad p_2 = 2.$$

Hence

$$g(x) = 2\int_0^{+\infty} t^{-s} \cos 2\pi xt\, dt = 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\Gamma(1-s)\, x^{s-1},$$

and

$$\lim_{N\to\infty}\left\{\sum_{n=1}^{N} n^{-s} - \frac{N^{1-s}}{1-s}\right\} = 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\Gamma(1-s) \lim_{N\to\infty}\left\{\sum_{n=1}^{N} n^{s-1} - \frac{N^s}{s}\right\}.$$

The expressions in brackets give the analytic continuation of $\zeta(s)$, and we have

$$\zeta(s) = 2(2\pi)^{s-1} \sin \tfrac{1}{2}s\pi\Gamma(1-s)\zeta(1-s),$$

the functional equation of the Riemann zeta-function.[12]

(ii). Consider

$$\varphi(z) = z^{-\frac{1}{2}} e^{\pi z/12} \sum_{n=0}^{\infty} p(n) e^{-2\pi nz} \qquad (z > 0),$$

---

[11] Cf. Erdelyi (1), Kober (1), Bochner (1), 37–38, and Titchmarsh (2), 64–65.

[12] Various methods of deriving this result from Poisson's formula have been given by other writers, but none is as direct as the above. Cf. Titchmarsh (2), 65.

where $p(n)$ is the number of unrestricted partitions of n. Using the generating function[13] of $p(n)$, we have

$$\varphi(z) = z^{-\frac{1}{4}} e^{\pi z/12} \prod_{n=1}^{\infty} (1 - e^{-2\pi n z})^{-1},$$

and

$$\log \varphi(z) = -\tfrac{1}{4} \log z + \frac{\pi z}{12} - \sum_{n=1}^{\infty} \log (1 - e^{-2\pi n z}).$$

Now put

$$f(x) = -\log (1 - e^{-2\pi x z})$$

in Theorem 1. The conditions are satisfied with $p = 2$, and

$$\begin{aligned} g(x) &= -2 \int_0^{\infty} \log (1 - e^{-2\pi t z}) \cos 2\pi x t \, dt \\ &= 2 \sum_{m=1}^{\infty} \frac{1}{m} \int_0^{\infty} e^{-2\pi m t z} \cos 2\pi x t \, dt \\ &= \frac{z}{\pi} \sum_{m=1}^{\infty} (m^2 z^2 + x^2)^{-1}. \end{aligned}$$

The term by term integration is justified by absolute convergence. If we apply Theorem 1 to this sum, we find that for $x > 0$

$$\begin{aligned} g(x) &= \frac{1}{x} \sum_{m=1}^{\infty} e^{-2\pi m x/z} + \frac{1}{2x} - \frac{z}{2\pi x^2} \\ &= \frac{1}{x} (e^{2\pi x/z} - 1)^{-1} + \frac{1}{2x} - \frac{z}{2\pi x^2}. \end{aligned}$$

Further

$$g(0) = \frac{z}{\pi} \sum_{m=1}^{\infty} \frac{1}{m^2 z^2} = \frac{\pi}{6z}.$$

Hence

$$\begin{aligned} &- \sum_{n=1}^{\infty} \log (1 - e^{-2\pi n z}) - \frac{\pi}{12z} \\ &\quad = \lim_{N \to \infty} \left[\sum_{n=1}^{\infty} \left\{ \frac{1}{n} (e^{2\pi n/z} - 1)^{-1} + \frac{1}{2n} - \frac{z}{2\pi n^2} \right\} \right. \\ &\qquad \left. - \int_0^N \left\{ \frac{1}{t} (e^{2\pi t/z} - 1)^{-1} + \frac{1}{2t} - \frac{z}{2\pi t^2} \right\} dt \right] \\ &\quad = \sum_{n=1}^{\infty} \frac{1}{n} (e^{2\pi n/z} - 1)^{-1} + \tfrac{1}{2}\gamma - \frac{\pi z}{12} \\ &\qquad - \lim_{N \to \infty} \left[\int_0^N \left\{ \frac{1}{t} (e^{2\pi t/z} - 1)^{-1} + \frac{1}{2t} - \frac{z}{2\pi t^2} \right\} dt - \tfrac{1}{2} \log N \right]. \end{aligned}$$

[13] Hardy and Wright (1), 272.

Now

$$\sum_{n=1}^{\infty} \frac{1}{n}(e^{2\pi n/z} - 1)^{-1} = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{m=1}^{\infty} e^{-2\pi mn/z}$$

$$= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{n} e^{-2\pi mn/z}$$

$$= -\sum_{m=1}^{\infty} \log (1 - e^{-2\pi m/z}),$$

and the above limit is equal to

$$\int_0^z \left\{\frac{1}{t}(e^{2\pi t/z} - 1)^{-1} + \frac{1}{2t} - \frac{z}{2\pi t^2}\right\} dt + \int_z^{\infty} \left\{\frac{1}{t}(e^{2\pi t/z} - 1)^{-1} - \frac{z}{2\pi t^2}\right\} dt$$

$$+ \frac{1}{2} \lim_{N \to \infty} \left[\int_z^N \frac{dt}{t} - \log N\right]$$

$$= \int_0^1 \left\{\frac{1}{u}(e^{2\pi u} - 1)^{-1} + \frac{1}{2u} - \frac{1}{2\pi u^2}\right\} du$$

$$+ \int_1^{\infty} \left\{\frac{1}{u}(e^{2\pi u} - 1)^{-1} - \frac{1}{2\pi u^2}\right\} du - \tfrac{1}{2} \log z$$

$$= K - \tfrac{1}{2} \log z,$$

where K is a constant. Hence

$$-\sum_{n=1}^{\infty} \log (1 - e^{-2\pi nz}) - \frac{\pi}{12z} = -\sum_{m=1}^{\infty} \log (1 - e^{-2\pi m/z}) - \frac{\pi z}{12} + \tfrac{1}{2}\gamma - K + \tfrac{1}{2} \log z.$$

That is

$$\log \varphi(z) = \log \varphi\left(\frac{1}{z}\right) + \tfrac{1}{2}\gamma - K.$$

Putting $z = 1$ it follows that $K = \frac{1}{2}\gamma$, and

$$\varphi(z) = \varphi\left(\frac{1}{z}\right).$$

Hence

$$z^{-\frac{1}{4}} e^{\pi z/12} \sum_{n=0}^{\infty} p(n) e^{-2\pi nz} = z^{\frac{1}{4}} e^{\pi/12z} \sum_{n=0}^{\infty} p(n) e^{-2\pi n/z}.$$

This is a well-known result in the theories of partitions and elliptic modular functions.

(iii). Put[14]

$$f(x) = (z^2 + x^2)^{-\frac{1}{2}}$$

[14] Cf. Watson (1).

then the conditions of Theorem 1 are satisfied with $p = 2$, and

$$g(x) = 2\int_0^{\infty} (z^2 + t^2)^{-\frac{1}{2}} \cos 2\pi xt\, dt = 2K_0(2\pi xz).$$

Hence

$$\begin{aligned} 2\sum_{n=1}^{\infty} K_0(2\pi nz) - \frac{1}{2z} &= \lim_{N\to\infty}\left[\sum_{n=1}^{N} (z^2+n^2)^{-\frac{1}{2}} - \int_0^N (z^2+t^2)^{-\frac{1}{2}}\, dt\right] \\ &= \lim_{N\to\infty}\left[\sum_{n=1}^{N}\left\{(z^2+n^2)^{-\frac{1}{2}} - \frac{1}{n}\right\} + \left\{\sum_{n=1}^{N}\frac{1}{n} - \log N\right\}\right. \\ &\qquad \left. + \left\{\log N - \int_0^{N/z} (1+u^2)^{-\frac{1}{2}}\, du\right\}\right] \\ &= \sum_{n=1}^{\infty}\left\{(z^2+n^2)^{-\frac{1}{2}} - \frac{1}{n}\right\} + \gamma - \lim_{N\to\infty}\left[\log\left\{\frac{1}{z} + \left(\frac{1}{z^2} + \frac{1}{N^2}\right)^{\frac{1}{2}}\right\}\right] \\ &= \sum_{n=1}^{\infty}\left\{(z^2+n^2)^{-\frac{1}{2}} - \frac{1}{n}\right\} + \gamma + \log \tfrac{1}{2}z. \end{aligned}$$

5. Series Involving Primitive Characters

Poisson's summation formula can be generalized by the introduction of the primitive characters as coefficients.[15] We require the following elementary properties of characters:[16]

(i). *If $\chi(n)$ is a character modulo k ($k > 1$) then*

$$\chi(m)\chi(n) = \chi(mn),$$
$$\chi(m + rk) = \chi(m)$$

where k, m, n, and r are integers. Further $\chi(n)$ vanishes if n is not prime to k.

(ii). *If $\chi(n)$ is a primitive character modulo k ($k > 1$) then*

$$\sum_{n=1}^{k} \chi(n) e^{2\pi mni/k} = \epsilon(\chi) k^{\frac{1}{2}} \bar{\chi}(m), \tag{5.1}$$

where

$$\epsilon(\chi) = k^{-\frac{1}{2}} \sum_{n=1}^{k} \chi(n) e^{2\pi ni/k}, \tag{5.2}$$

and $\bar{\chi}(m)$ is the conjugate of $\chi(m)$.

Now $\chi(-1) = \pm 1$, since

$$\{\chi(-1)\}^2 = \chi(1) = 1.$$

---

[15] Some particular cases of this type of generalization have been given by Ramanujan. Cf. Titchmarsh (2), 82.

[16] Landau (1), chapters XXII and XXX. The notation for $\epsilon(\chi)$ used here differs from Landau's by a factor $-i$ in the case $\chi(-1) = -1$.

Suppose that $\chi(-1) = 1$; then

$$\sum_{n=1}^{k} \chi(n) \sin \frac{2\pi mn}{k} = 0,$$

since

$$\chi(k - r) = \chi(-r) = \chi(-1)\chi(r) = \chi(r),$$

and consequently the terms $n = r$ and $n = k - r$ cancel. Also the term $n = \frac{1}{2}k$, if it occurs, is zero since $\chi(\frac{1}{2}k) = 0$. Hence (5.1) becomes

$$\sum_{n=1}^{k} \chi(n) \cos \frac{2\pi mn}{k} = \epsilon(\chi) k^{\frac{1}{2}} \bar{\chi}(m). \tag{5.3}$$

Now if we put

$$f(x) = \begin{cases} \cos \dfrac{2\pi mx}{k} & (-y < x < y), \\ 0 & (|x| > y) \end{cases}$$

in (1.2), we find that

$$\sum_{-y \leqq n \leqq y}{}' \cos \frac{2\pi mn}{k} = \frac{k}{\pi} \sum_{n=-\infty}^{\infty} \frac{1}{nk + m} \sin \frac{2\pi y}{k} (nk + m),$$

where the prime indicates that the terms $n = \pm y$ are to be halved if y is an integer. If we multiply this result by $\chi(m)$, and sum from $m = 1$ to $m = k$, we have

$$\sum_{m=1}^{k} \chi(m) \sum_{-y \leqq n \leqq y}{}' \cos \frac{2\pi mn}{k} = \frac{k}{\pi} \sum_{m=1}^{k} \chi(m) \sum_{n=-\infty}^{\infty} \frac{1}{nk + m} \sin \frac{2\pi y}{k} (nk + m).$$

By (5.3) this is

$$\begin{aligned} \epsilon(\chi) k^{\frac{1}{2}} \sum_{-y \leqq n \leqq y}{}' \bar{\chi}(n) &= \frac{k}{\pi} \sum_{m=1}^{k} \sum_{n=-\infty}^{\infty} \frac{\chi(nk + m)}{nk + m} \sin \frac{2\pi y}{k} (nk + m) \\ &= \frac{k}{\pi} \sum_{n=-\infty}^{\infty} \frac{\chi(n)}{n} \sin \frac{2\pi ny}{k}. \end{aligned}$$

Hence

$$\epsilon(\chi) k^{\frac{1}{2}} \sum_{1 \leqq n \leqq y}{}' \bar{\chi}(n) = \frac{k}{\pi} \sum_{n=1}^{\infty} \frac{\chi(n)}{n} \sin \frac{2\pi ny}{k}.$$

Now let us consider the integral

$$2\pi k^{-\frac{1}{2}} \int_0^{+\infty} t^{-1} \left\{ \sum_{1 \leqq n \leqq t}{}' \chi(n) \right\} \left(\frac{xt}{k} \right)^{\frac{1}{2}} J_{\frac{1}{2}} \left(\frac{2\pi xt}{k} \right) dt. \tag{5.4}$$

We have

$$\frac{2\pi x^{\frac{1}{2}}}{k}\int_0^N \Big\{\sum_{1\leqq n\leqq t}{}' \chi(n)\Big\}\, t^{-\frac{1}{2}} J_{\frac{1}{2}}\left(\frac{2\pi xt}{k}\right) dt$$

$$= \frac{2\pi x^{\frac{1}{2}}}{k}\sum_{n=1}^{N-1}\{\chi(1)+\chi(2)+\cdots+\chi(n)\}\int_n^{n+1} t^{-\frac{1}{2}} J_{\frac{1}{2}}\left(\frac{2\pi xt}{k}\right) dt$$

$$= \frac{k^{\frac{1}{2}}}{\pi x}\sum_{n=1}^{N-1}\{\chi(1)+\chi(2)+\cdots+\chi(n)\}\left\{\frac{1}{n}\sin\frac{2\pi nx}{k} - \frac{1}{n+1}\sin\frac{2\pi(n+1)x}{k}\right\}$$

$$= \frac{k^{\frac{1}{2}}}{\pi x}\sum_{n=1}^{N}\frac{\chi(n)}{n}\sin\frac{2\pi nx}{k} - \frac{k}{\pi Nx}\sin\frac{2\pi Nx}{k}\sum_{n=1}^{N}\chi(n).$$

If we make N tend to infinity the integral (5.4) becomes

$$\frac{k^{\frac{1}{2}}}{\pi x}\sum_{n=1}^{\infty}\frac{\chi(n)}{n}\sin\frac{2\pi nx}{k} = \epsilon(\chi)x^{-1}\Big\{\sum_{1\leqq n\leqq x}{}'\bar{\chi}(n)\Big\}.$$

That is the functions

$$x^{-1}\Big\{\sum_{1\leqq n\leqq x}{}'\chi(n)\Big\}, \qquad \epsilon(\chi)x^{-1}\Big\{\sum_{1\leqq n\leqq x}{}'\bar{\chi}(n)\Big\}$$

are transforms with respect to the kernel

$$\frac{2\pi x^{\frac{1}{2}}}{k}J_{\frac{1}{2}}\left(\frac{2\pi x}{k}\right).$$

If we now apply the methods of section 3 we can obtain a result analogous to Theorem 1. We can also obtain a result corresponding to the case $\chi(-1) = -1$. We have:

THEOREM 4. *If $\chi(n)$ is a primitive character modulo k $(k > 1)$, $f(x)$ is an integral, tends to zero at infinity, $xf'(x)$ belongs to $L^p(0, \infty)$, $(1 < p \leqq 2)$, and either* (i) $\chi(-1) = 1$,

$$g(x) = 2\epsilon(\chi)k^{-\frac{1}{2}}\int_0^{\infty}\cos\frac{2\pi xt}{k}f(t)\,dt,$$

or (ii) $\chi(-1) = -1$,

$$g(x) = -2i\epsilon(\chi)k^{-\frac{1}{2}}\int_0^{\infty}\sin\frac{2\pi xt}{k}f(t)\,dt,$$

then

$$\sum_{n=1}^{\infty}\chi(n)f(n) = \sum_{n=1}^{\infty}\bar{\chi}(n)g(n).$$

For real primitive characters $\bar{\chi}(n) = \chi(n)$, and it follows on substituting $f(x) = e^{-\pi x^2/k} = g(x)$ in case (i) that $\epsilon(\chi) = 1$. Similarly in case (ii) $f(x) = xe^{-\pi x^2/k} = g(x)$ gives $-i\epsilon(\chi) = 1$, and we have:

THEOREM 5. *If, in addition to the assumptions of Theorem* 4, $\chi(n)$ *is a real character, and either*

(i) $\chi(-1) = 1$,

$$g(x) = 2k^{-\frac{1}{2}} \int_0^{+\infty} \cos \frac{2\pi xt}{k} f(t)\, dt,$$

or (ii) $\chi(-1) = -1$,

$$g(x) = 2k^{-\frac{1}{2}} \int_0^{+\infty} \sin \frac{2\pi xt}{k} f(t)\, dt,$$

then

$$\sum_{n=1}^{\infty} \chi(n) f(n) = \sum_{n=1}^{\infty} \chi(n) g(n).$$

PRINCETON, NEW JERSEY.

BIBLIOGRAPHY

S. Bochner (1), Vorlesungen über Fouriersche Integrale, (Leipzig 1932).

I. W. Busbridge (1), A theory of general transforms for functions of the class $L^p(0, \infty)$ $(1 < p \leqq 2)$, Quart. J. of Math. (Oxford), 9 (1938), 148–160.

R. Courant and D. Hilbert (1), Methoden der mathematischen Physik I, (Berlin 1931).

A. Erdelyi (1), Gewisse Reihentransformationen, die mit der linearen Transformationsformel der Theta-funktion zusammenhangen, Compositio Math. 4 (1937), 406–423.

W. L. Ferrar (1), Summation formulae and their relation to Dirichlet's series II, Compositio Math. 4 (1937), 394–405.

A. P. Guinand (1), Summation formulae and self-reciprocal functions, Quart. J. of Math. (Oxford), 9 (1938), 53–67.

G. H. Hardy and E. M. Wright (1), Introduction to the theory of numbers, (Oxford 1938).

H. Kober (1), Transformationsformeln gewisser Besselscher Reihen. Beziehungen zu Zeta-funktionen, Math. Zeitschrift 39 (1934), 609–624. (2), Hankelsche Transformationen, Quart. J. of Math. (Oxford), 8 (1937), 186–199.

L. Kronecker (1), Bemerkungen über die Darstellung von Reihen durch Integrale, Journal für Math. 105 (1889), 157–159 and 345–354.

E. Landau (1), Handbuch der Lehre von der Verteilung der Primzahlen, (Leipzig 1909).

G. Landsberg (1), Zur Theorie Gausschen Summen und der linearen Transformation der Theta-funktionen, Journal für Math. 111 (1893), 234–253.

E. C. Titchmarsh (1), Theory of functions, (Oxford 1939). (2), Introduction to the theory of Fourier integrals, (Oxford 1937).

G. N. Watson (1), Some self-reciprocal functions, Quart. J. of Math. (Oxford), 2 (1931), 298–309.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON THE REARRANGEMENT OF CONDITIONALLY CONVERGENT SERIES

BY C. T. RAJAGOPAL

(Received June 25, 1940)

This note is concerned with an application of the idea which finds expression in

LITTLEWOOD'S THEOREM [2].[1] *If $F(x)$ is positive monotone decreasing and $\{D_n\}$ is a sequence such that*

$$\left.\begin{aligned} 0 = D_0 < D_1 < D_2 < \cdots < D_n \to \infty, \\ D_n - D_{n-1} = O(1), \end{aligned}\right\} \quad n \to \infty,$$

then $\sum^{\infty} (D_n - D_{n-1})F(D_n)$ converges or diverges with $\int^{\infty} F(x)\,dx$.

In an earlier note [6], I have shown that the idea behind the theorem—that of a condensation effected on series by integrating a suitable function over intervals of the type (D_{n-1}, D_n)—supplies the basis for a unified theory of convergence criteria for series of positive terms, more comprehensive than Pringsheim's [3] in certain respects. In another note [7], I have shown that the idea suggests elegant proofs of certain classic theorems of Pringsheim [4]. I prove here that the same idea can be used to generalize the enunciations and simplify the proofs of Pringsheim's theorems concerning the rearrangement of conditionally convergent series [5].[2]

1. Notation

$S \equiv \sum_{n=1}^{\infty} (-1)^{n-1}a_n (a_n > 0)$ is a conditionally convergent series, $s_n = \sum_{\nu=1}^{n} (-1)^{\nu-1}a_\nu$. $\sum_{n=1}^{\infty} (-1)^{n-1}a_n'$ is a rearrangement of s in which $s_{p+q}' = \sum_{\nu=1}^{p+q} (-1)^{\nu-1}a_\nu'$ contains the first p positive terms and the first q negative terms of s.

1.1. The theorems which follow are typical of the results obtained by Pringsheim. It is supposed in their enunciations that $a_n \equiv a(n)$, where $a(x)$ is a differentiable function of x, and that limiting relations exist, as specified, between p and q.

THEOREM A. *If, in $S \equiv \sum_1^{\infty} (-1)^{n-1}a_n$, a_n^{-1} and na_n tend steadily to infinity with n, then $\sum_1^{\infty} (-1)^{n-1}a_n'$ has a sum s' given by*

$$s' = s + \lim_{p>q\to\infty} (p - q)a_{2q}.$$

[1] The numbers in [ ] denote the references at the end of this note. A slight generalization of Littlewood's theorem, which dispenses with the condition: $D_n - D_{n-1} = O(1)$, is given in de la Vallée-Poussin, *Cours d'analyse*, 1 (Louvain-Paris, 1926), 398. A generalization in a different direction is given in Rajagopal [8], §2.

[2] All the references to Pringsheim in the sequel are to [5].

THEOREM B. *If, in s,* $\lim_{n\to\infty} na_n$ *is finite, equal to g (say), then*

$$s' = s + \frac{g}{2} \lim_{p,q\to\infty} (\log 2p - \log 2q).$$

THEOREM B′. *If, in s,* $\lim_{n\to\infty} n \cdot l_1 n \cdot l_2 n \cdots l_{x-1} n \cdot a_n = g$ *(where* $l_1 n = \log n$, $l_2 n = \log\log n, \cdots$*), then*

$$s' = s + \frac{g}{2} \lim_{p,q\to\infty} (l_x 2p - l_x 2q).$$

THEOREM C. *If, in s,* na_1 *tends steadily to the limit zero and if* $(p-q)a_{2q} = O(1)$ $(p > q \to \infty)$, *then* $s' = s$.

Theorems A, B, C, B′ are suggested by (26), (30), (31), (34) respectively in Pringsheim's paper. The restriction on a_n in all the theorems is of the form: $\lim_{n\to\infty} f(n)\cdot a_n$ is finite, where $f(x)$ is a positive differentiable function for all large x. The distinction between the theorems arises from the fact that in Theorem A, $\lim f(n)/n = 0$; in Theorem B, $\lim f(n)/n$ is finite and non-zero; in Theorems B′ and C, $\lim f(n)/n = \infty$. A question, suggested by Littlewood's theorem, is whether we cannot with advantage replace $f(n)$ by $f(D_n)$ in the condition imposed on a_n. An attempt at an answer is contained in Theorems 1 and 1a which are found to cover all the cases discussed by Pringsheim.

In all the theorems and the lemmas proved below it is tacitly assumed that $0 < d_n \equiv D_n - D_{n-1} < K$ *(fixed).*

2

THEOREM 1. *If* $f(x)$ *is a positive function with a continuous derivative* $f'(x)$ *such that* $\frac{f'(x)}{f(x)} = o(1)$ $(x \to \infty)$, *and if* $\lim_{n\to\infty} \frac{a_n \cdot f(D_n)}{D_n - D_{n-2}} = g$, *then* $\sum_1^\infty (-1)^{n-1} a_n'$ *oscillates between* $s + g \varliminf_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$, *provided* $\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$ *is bounded.*

PROOF:

$$\frac{D_\nu - D_{\nu-2}}{f(D_\nu)} = \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} + \int_{D_{\nu-2}}^{D_\nu} \left[\frac{1}{f(D_\nu)} - \frac{1}{f(x)}\right] d(x - D_{\nu-2}).$$

Integrating by parts the last term of this identity, we find

$$\begin{aligned}\frac{D_\nu - D_{\nu-2}}{f(D_\nu)} &= \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} + \int_{D_{\nu-2}}^{D_\nu} (x - D_{\nu-2})\, d\left[\frac{1}{f(x)}\right] \\ &= \int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} + o(1)\int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} \qquad (\nu \to \infty);\end{aligned}$$

whence, given $\epsilon > 0$ arbitrarily small, we can find n_0 such that

$$(1) \qquad (g-\epsilon)\int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} < a_\nu < (g+\epsilon)\int_{D_{\nu-2}}^{D_\nu} \frac{dx}{f(x)} \qquad (\nu \geqq n_0).$$

If for the sake of definiteness, we suppose $p > q$,

$$(2) \qquad \sum_{\nu=1}^{p+q} (-1)^{\nu-1} a'_\nu = \sum_{\nu=1}^{2q} (-1)^{\nu-1} a_\nu + (a_{2q+1} + a_{2q+3} + \cdots + a_{2p-1}).$$

Hence, if $2q + 1 \geqq n_0$, summing (1) for $\nu = 2q + 1, 2q + 3, \cdots 2p - 1$, we obtain from (2),

$$(3) \qquad s'_{p+q} \begin{cases} < s_{2q} + (g + \epsilon) \int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)}, \\ > s_{2q} + (g - \epsilon) \int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)}, \end{cases}$$

where we can write

$$(4) \qquad \int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)} = \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} + \int_{D_{2q-1}}^{D_{2p}} \frac{dx}{f(x)} - \int_{D_{2p-1}}^{D_{2p}} \frac{dx}{f(x)}$$

and show that the second and third terms of the right hand member are $o(1)$ as $p, q \to \infty$.

First, when $0 < v - u = O(1)$ $(u \to \infty)$,

$$\log \frac{f(v)}{f(u)} = \int_u^v \frac{f'(x)}{f(x)} dx = o(1) \qquad (u \to \infty).$$

Or,

$$(5) \qquad f(v) \sim f(u).$$

Next,

$$\int_{D_{2n-1}}^{D_{2n}} \frac{dx}{f(x)} = \frac{D_{2n} - D_{2n-1}}{f\{D_{2n-1} + \theta(D_{2n} - D_{2n-1})\}} \qquad (0 < \theta < 1)$$

$$\sim \frac{D_{2n} - D_{2n-1}}{f(D_{2n})} \qquad (n \to \infty),$$

in virtue of (5).

Also,

$$0 < \frac{D_{2n} - D_{2n-1}}{f(D_{2n})} < \frac{D_{2n} - D_{2n-2}}{f(D_{2n})} = o(1) \qquad (n \to \infty),$$

in consequence of the restriction $\frac{D_N - D_{N-2}}{f(D_N)} = o(1)$ $(N \to \infty)$ which results from $a_N = o(1)$ and $\frac{a_N}{(D_N - D_{N-2})/f(D_N)} \sim g$; whence

$$(6) \qquad \int_{D_{2n-1}}^{D_{2n}} \frac{dx}{f(x)} = o(1) \qquad (n \to \infty).$$

We can now complete the proof by using (6) in (4) and (4) in (3).

In certain applications it is useful to have the following complement to Theorem 1.

THEOREM 1a. *If, in Theorem* 1, *we replace the restriction relating to* $f'(x)$ *by the assumption that* $f(x)$ *tends steadily to* ∞ *with* x, *then either* $s'_{p+q} - s_{2q}$ $(p > q)$, *or* $s'_{p+q} - s_{2p}$ $(q > p)$, *has an asymptotic representation which oscillates between* $g \varlimsup_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$, *provided* $\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$ *is bounded.*

PROOF: Supposing (as before) $p > q$,

$$\frac{D_{2q+1} - D_{2q-1}}{f(D_{2q+1})} + \frac{D_{2q+3} - D_{2q+1}}{f(D_{2q+3})} + \cdots + \frac{D_{2p-1} - D_{2p-3}}{f(D_{2p-1})} = \int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)} + o(1) \qquad (q \to \infty).$$

As in (4), we can split $\int_{D_{2q-1}}^{D_{2p-1}} \frac{dx}{f(x)}$ into three terms of which the last two are $o(1)$ as $p, q \to \infty$, in virtue of the condition: $f(x) \to \infty$ with x.

Hence, as $q \to \infty$,

$$a_{2q+1} + a_{2q+3} + \cdots + a_{2p-1} \sim g\left[\frac{D_{2q+1} - D_{2q-1}}{f(D_{2q+1})} + \cdots + \frac{D_{2p-1} - D_{2p-3}}{f(D_{2p-1})}\right]$$

$$= g \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} + o(1)$$

which oscillates between $g \varlimsup \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)}$.

REMARK. The argument employed to prove Theorems 1 and 1a shows that, under the conditions of these theorems, the sum s' of the rearranged series is given by

$$s' = s + g \lim \int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} \tag{7}$$

provided this limit exists and is finite. If the limit is infinite, then $\sum_1^\infty (-1)^{n-1} a'_n$ diverges.

2.1. We may consider, in the light of this remark, series of the type in Pringsheim's paper, in forms slightly more general than his. The generalized series are all of the kind whose conditional convergence we can establish by

LEMMA 1. *If, as* $x \to \infty$, $f(x)$ *decreases steadily to the limit zero and* $\int^\infty F(x)\,dx$ *is divergent, then* $\sum^\infty (-1)^{n-1} F(D_n)$ *is conditionally convergent.*

PROOF: The convergence of the series in question follows from Leibnitz's rule for alternating series.

To prove that the convergence is not absolute, we observe that $KF(D_n) > (D_n - D_{n-1})F(D_n)$ which, by Littlewood's theorem, is the general term of a divergent series.

Examples of the lemma are furnished by $\sum^{\infty}(-1)^{n-1}a_n$, where a_n may be any one of the functions:

$$\frac{1}{D_n^{\mu}}\ (0<\mu\leqq 1),\quad \frac{1}{\log D_n},\quad \frac{\log D_n}{D_n},\quad \frac{1}{D_n\cdot\log D_n},\quad \frac{1}{AD_n+BD_n^{\mu}}\ \begin{pmatrix}A,B>0;\\ 1>\mu>0\end{pmatrix}.$$

A second proposition from which we can infer the conditional convergence of a certain type of alternating series is

LEMMA 2. *If* $u_n>0$ *is such that*

$$\frac{u_n/d_n}{u_{n-1}/d_{n-1}}=1-\mu\frac{d_n}{D_n}-O\left(\frac{d_n}{D_n^{\lambda}}\right)\qquad(\lambda>1),\tag{8}$$

then (i) $\sum^{\infty}u_n$ *is convergent or divergent according as* $\mu>0$, *or* $\leqq 1$; (ii) $\sum^{\infty}(-1)^{n-1}u_n/d_n$ *is convergent for* $\mu>0$, (iii) *the convergence in* (ii) *is conditional for* $0<\mu\leqq 1$.

This result is at once an extension and an amplification of Gauss's test.

(i) and (ii) are immediate consequences of the relation: $\frac{u_n}{d_n}\sim cD_n^{-\mu}$, equivalent to (8)[3].

(iii) follows from the fact that $u_n/d_n>u_n/k$ which, by (i), is the general term of a divergent series when $\mu\leqq 1$.

As an application of Lemma 2, consider the case:

$$u_n=d_n\frac{\prod_{\nu=1}^{n}(\alpha d_\nu+D_\nu)(\beta d_\nu+D_\nu)}{\prod_{\nu=1}^{n}(\gamma d_\nu+D_\nu)(\delta d_\nu+D_\nu)}$$

where α, β, γ, δ are real and the two latter $\neq -D_n/d_n$ $(n=1,2,\cdots)$. It is clear that $\sum u_n$ reduces to the hypergeometric type when $d_n=1$, $\delta=0$. Further,

$$\frac{u_n/d_n}{u_{n-1}/d_{n-1}}=1-(\gamma+\delta-\alpha-\beta)\frac{d_n}{D_n}-O\left(\frac{d_n}{D_n^2}\right).$$

The terms of $\sum^{\infty}u_n$ being ultimately of the same sign and the terms of $\sum^{\infty}(-1)^{n-1}u_n/d_n$ alternately positive and negative, Lemma 2 shows that

$$\sum^{\infty}(-1)^{n-1}a_n\equiv\sum^{\infty}(-1)^{n-1}\frac{\prod_{\nu=1}^{n}(\alpha d_\nu+D_\nu)(\beta d_\nu+D_\nu)}{\prod_{\nu=1}^{n}(\gamma d_\nu+D_\nu)(\delta d_\nu-D_\nu)}\tag{9}$$

is conditionally convergent for $0<\gamma+\delta-\alpha-\beta\leqq 1$.

REMARK. a_n in (9) can be related to a generalization of Gauss's function:

$$\left.\begin{aligned}\Pi(z)&=\lim_{n\to\infty}\frac{n^z}{(1+z)\left(1+\frac{z}{2}\right)\cdots\left(1+\frac{z}{n}\right)}\\ &=\Gamma(1+z)\end{aligned}\right\}\ (z\neq -1,-2,\cdots)$$

[3] Lemma 2 (i) is, in fact, a particular case of a theorem first proved by me in *Bull. Amer. Math. Soc.* 43 (1937), 411. But the proof of the theorem given there is not applicable to Lemma 2 (ii).

whose reciprocal is, in Weierstrass's form:

$$\frac{1}{\Pi(z)} = e^{Cz} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) e^{-z/n},$$

C being Euler's constant.

The generalization of $\Pi(z)$ referred to is due to Cesàro[4]. It is given by

$$\Pi_{(d_n)}(z) = \lim_{n\to\infty} \frac{D_n^z}{\left(1 + z\frac{d_1}{D_1}\right)\left(1 + z\frac{d_2}{D_2}\right)\cdots\left(1 + z\frac{d_n}{D_n}\right)} \tag{10}$$

$$\left(z \neq \frac{-D_1}{d_1}, \frac{-D_2}{d_2}, \cdots\right).$$

Its reciprocal can be expressed in the form:

$$\frac{1}{\Pi_{(d_n)}(z)} = e^{C_{(d_n)}\cdot z} \prod_{n=1}^{\infty} \left(1 + z\frac{d_n}{D_n}\right) e^{-zd_n/D_n}$$

where $C_{(d_n)} = \lim_{n\to\infty} \left(\sum_{\nu=1}^{n} \frac{d_\nu}{D_\nu} - \log D_n\right)$ is a constant with some of the elementary properties and application of C.[5]

(10) *shows that* a_n *in* (9) *is subject to the condition*

$$\lim_{n\to\infty} a_n \cdot D_n^{\gamma+\delta-\alpha-\beta} = \frac{\Pi_{(d_n)}(\gamma)\cdot\Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha)\cdot\Pi_{(d_n)}(\beta)}. \tag{11}$$

This fact will be used later (§3.1).

2.2. Taking the series in §2.1, we can find out how they are affected by particular rearrangements, supposing that in all of them appearing as examples, $d_n \sim k > 0$ $(n \to \infty)$.

EXAMPLES. 1) Let $\sum_1^\infty (-1)^{n-1} a_n \equiv \sum_1^\infty (-1)^{n-1}/D_n \cdot \log D_n$. Taking $f(x) = x\cdot\log x$ in Theorem 1, we find $g = 1/2k$. Hence, by (7),

$$s' - s = \frac{1}{2k} \lim_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{x\cdot\log x} = \frac{1}{2k} \lim \log\left(\frac{\log D_{2p}}{\log D_{2q}}\right)$$

provided this limit exists.

In particular, if $0 < \underline{\lim}\ (D_{2p}/D_{2q}) < \infty$, there is no alternation in the sum of the series, due to the rearrangement.

Next, if $D_{2p} \sim D_{2q}^{\lambda}$ $(\lambda > 1)$, the alteration is $(\log \lambda)/2k$.

Finally, if $D_{2p} \sim \lambda^{D_{2q}}$ $(\lambda > 1)$, the rearranged series diverges to ∞.

2) Let $\sum_1^\infty (-1)^{n-1} a_n \equiv \sum_1^\infty (-1)^{n-1}/AD_n + BD_n^{\mu}$ $(A, B > 0; 1 > \mu > 0)$.

[4] The function is mentioned in E. T. Whittaker, *Modern Analysis* (Cambridge, 1902), 195. The notation in Whittaker is, however, different from the one employed here.

[5] E.g., Rajagopal [8], §3.

With $f(x) = Ax + Bx^{\mu}$, $g = 1/2k$ so that

$$s' - s = \frac{1}{2k} \lim_{p,q\to\infty} \int_{D_{2q}}^{D_{2p}} \frac{dx}{Ax + Bx^{\mu}} = \frac{1}{2kA(1-\mu)} \lim \log \frac{AD_{2p}^{1-\mu} + B}{AD_{2q}^{1-\mu} + B}$$

if this limit exists.

Thus, if $D_{2p} \sim bD_{2q}$, the alteration in the sum of the series is $(\log b)/2kA$.

3) Let $\sum_1^{\infty} (-1)^{n-1} a_n \equiv \sum_1^{\infty} (-1)^{n-1} \frac{\log D_n}{D_n}$. If $f(x) = \frac{x}{\log x}$, $g = \frac{1}{2k}$ and so

$$s' - s = \frac{1}{2k} \lim_{p,q\to\infty} \left[\frac{(\log x)^2}{2}\right]_{D_{2q}}^{D_{2p}}$$

if the limit exists.

Suppose that as $p > q \to \infty$,

$$\frac{D_{2p}}{D_{2q}} = 1 + \frac{a + o(1)}{\log D_{2q}}.$$

Then a simple calculation shows that the alteration in the sum of the series is $a/2k$.

3

These examples typify, in the order of their discussion, the three cases: $\lim_{x\to\infty} f(x)/x = \infty$, $0 < \lim f(x)/x < \infty$, $\lim f(x)/x = 0$. The function $f(x)$ which figures in them is such that $\int \frac{dx}{f(x)}$ can be readily evaluated. Under certain conditions, it is possible to find $s' - s$ without evaluating the integral. The conditions are set forth below in three theorems which reflect the differences between the three types of restriction on $f(x)$.

THEOREM 2. *If* $\lim_{n\to\infty} \frac{a_n \cdot D_n}{D_n - D_{n-2}} = g$, *then* $s' - s = g \lim_{p,q\to\infty} \log\left(\frac{D_{2p}}{D_{2q}}\right)$ *provided the limit exists.*

PROOF. Take $f(x) = x$ in (7)

COROLLARY. *Setting* $D_n = n, l_1 n, l_2 n \cdots$ *successively, we obtain Theorems* B *and* B′.

THEOREM 3. *If* $f(x)/x$ *tends steadily to* ∞ *with* x,

$$\lim_{n\to\infty} \frac{a_n \cdot f(D_n)}{D_n - D_{n-2}} = g, \qquad \frac{D_{2p} - D_{2q}}{f(D_{2q})} = O(1) \qquad (p > q \to \infty),$$

then s *is unaltered by the rearrangement.*[6]

[6] It is obvious that Theorems 3, 4 and all deductions therefrom can be suitably restated in the case $q > p$.

PROOF. Setting $f(x) = x \cdot \varphi(x)$, we have

$$\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} = \frac{1}{\varphi\{D_{2q} + \theta(D_{2p} - D_{2q})\}} \int_{D_{2q}}^{D_{2p}} \frac{dx}{x} \qquad (0 < \theta < 1)$$

$$= \frac{\log\left\{1 + \dfrac{D_{2p} - D_{2q}}{f(D_{2q})}\varphi(D_{2q})\right\}}{\varphi\{D_{2q} + \theta(D_{2p} - D_{2q})\}} = o(1) \qquad (q \to \infty).$$

The desired result now follows from Theorem 1a.

COROLLARY. *The particular case* $D_n = n$ *contains Theorem* C.

More generally, we may take $D_n = l_1 n, l_2 n, \cdots$

THEOREM 4. *If, in Theorem* 1, $f(x)$ *satisfies the additional condition*

$$\lim_{x \to \infty} f'(x) = 0,$$

then $s' - s = 0$ *or* ag *or* ∞ *according as* $\lim\limits_{p > q \to \infty} \dfrac{D_{2p} - D_{2q}}{f(D_{2q})} = 0$ *or* a *or* ∞.

PROOF. Since

$$\int_{D_{2q}}^{D_{2p}} \frac{dx}{f(x)} = \frac{D_{2p} - D_{2q}}{f(D_{2q}) + \theta(D_{2p} - D_{2q}) f'\{D_{2q} + \eta(D_{2p} - D_{2q})\}} \qquad (0 < \eta < \theta < 1),$$

the result sought is an immediate consequence of (7).

REMARK. To bring out the relation between Theorems 3 and 4, we need only vary the form of the hypotheses in the latter to the extent of replacing the restriction on $f'(x)$ by the conditions: $f(x)$ tends steadily to ∞ with x and $f(x)/x$ tends steadily to zero. These conditions imply that $\lim f'(x) = 0$ and therefore, our previous proof of Theorem 4 is still valid.

COROLLARY 1. *When* $D_n = n$, $a_n = 1/f(n)$ *in the special form of Theorem* 4 *noticed above, we have Theorem* A.

COROLLARY 2. Let $D_n = l_x n$ $(x \geqq 1)$,

$$f(x) = (l_k x)^{m_k} (l_{k+1} x)^{m_{k+1}} \cdots (l_\lambda x)^{m_\lambda} \xi(x) \equiv \psi(x) \text{ (say)}$$

where $k \geqq 1$, the ms are positive or negative numbers, $\xi(x)$ has a continuous derivative and is such that $\xi(x) \to \infty$, $\xi'(x) = 0(1/x)$ $(x \to \infty)$. Theorem 4 then gives the following equivalent of (38a) and (38b) in Pringsheim's paper:

If $\lim_{n \to \infty} l_0 n \cdot l_1 n \cdot l_2 n \cdots l_{x-1} n \cdot \psi(l_x n) \cdot a_n = g$ $(l_0 n = n, x \geqq 1)$, *then* $s' - s = 0$ *or* $ag/2$ *or* ∞ *according as* $\lim\limits_{p > q \to \infty} \dfrac{l_x(2p) - l_x(2q)}{\psi(l_x 2q)} = 0$ *or* a *or* ∞.

COROLLARY 3. Let $D_n = l_x n$ $(x \geqq 1)$, $f(x) = x^{1-\rho} \cdot \psi(x)$ $(\rho > 0)$. We have then the result:

If $\lim_{n \to \infty} l_0 n \cdot l_1 n \cdot l_2 n \cdots l_{x-1} n \cdot (l_x n)^{1-\rho} \cdot \psi(l_x n) \cdot a_n = g$ $(l_0 n = n, x \geqq 1)$, *then* $s' - s = 0$ *or* $ag/2$ *or* ∞ *according as* $\lim\limits_{p > q \to \infty} \dfrac{l_x(2p) - l_x(2q)}{(l_x 2q)^{1-\rho} \cdot \psi(l_x 2q)} = 0$ *or* a *or* ∞.

This is equivalent to (44) in Pringsheim's paper.

NOTE. Theorem 4 thus includes two of the limiting cases discussed by Pringsheim: $\lim na_n = \infty$ as in Corollary 1, $\lim na_n = 0$ as in Corollaries 2 and 3.

3.1. I conclude with some applications of the theorems in §3. It is assumed in all these applications that $d_n \sim k$ as in §2.2.

1) If the series $\sum_1^\infty (-1)^{n-1}/D_n \cdot \log D_n$ is rearranged so that $D_{2p} = O(D_{2q} \cdot \log D_{2q})$ $(p > q \to \infty)$, then its sum is unaffected by the rearrangement (Theorem 3).

2) If the series $\sum_1^\infty (-1)^{n-1}/\log D_n$ is rearranged so that $D_{2p} - D_{2q} \sim a \cdot \log D_{2q}$ $(p > q \to \infty)$, then the alteration in its sum is $a/2k$ (Theorem 4).

3) Consider the series in (9), introducing the additional restriction $d_n \sim k$. (11) then gives

$$\lim \frac{a_n \cdot D_n^{\gamma+\delta-\alpha-\beta}}{D_n - D_{n-2}} = \frac{1}{2k} \frac{\Pi_{(d_n)}(\gamma) \cdot \Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha) \cdot \Pi_{(d_n)}(\beta)};$$

so that if $0 < \gamma + \delta - \alpha - \beta < 1$, a_n is subject to the hypothesis in Theorem 4, with $f(x) = x^{\gamma+\delta-\alpha-\beta}$; and if $\gamma + \delta - \alpha - \beta = 1$, a_n is as in Theorem 2. Hence we are led to formulate:

Suppose that for the series in (9) $d_n \sim k$.

(i) *If* $0 < \gamma + \delta - \alpha - \beta < 1$ *and if the series is rearranged so that* $D_{2p} - D_{2q} \sim aD_{2q}^{\gamma+\delta-\alpha-\beta}$ $(p > q \to \infty)$, *then the alteration in the sum of the series due to the rearrangement is*

$$\frac{a}{2k} \frac{\Pi_{(d_n)}(\gamma) \cdot \Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha) \cdot \Pi_{(d_n)}(\beta)}.$$

(ii) *If* $\gamma + \delta - \alpha - \beta = 1$ *and if in the rearrangement* $D_{2p} \sim bD_{2q}$ $(p > q \to \infty)$, *then the alteration is* $\frac{\log b}{2k} \cdot \frac{\Pi_{(d_n)}(\gamma) \cdot \Pi_{(d_n)}(\delta)}{\Pi_{(d_n)}(\alpha) \cdot \Pi_{(d_n)}(\beta)}$.

In the particular case $d_n = 1$, we have:

Suppose that the series

$$\sum_1^\infty (-1)^{n-1} \frac{(\alpha+1)(\alpha+2)\cdots(\alpha+n)\cdot(\beta+1)(\beta+2)\cdots(\beta+n)}{(\gamma+1)(\gamma+2)\cdots(\gamma+n)\cdot(\delta+1)(\delta+2)\cdots(\delta+n)}$$

where $\alpha, \beta, \gamma, \delta$ *are real and* $\gamma, \delta \neq -1, -2, \cdots$, *is rearranged* (i) *in the case* $0 < \gamma + \delta - \alpha - \beta < 1$, *so that* $p - q \sim aq^{\gamma+\delta-\alpha-\beta}$ $(p > q \to \infty)$, (ii) *in the case* $\gamma + \delta - \alpha - \beta = 1$, *so that* $p \sim bq$ $(p, q \to \infty)$. *Then the alterations in the sum of the series in* (i) *and* (ii) *are respectively* $\frac{a}{2^{\gamma+\delta-\alpha-\beta}} \frac{\Gamma(1+\gamma)\cdot\Gamma(1+\delta)}{\Gamma(1+\alpha)\cdot\Gamma(1+\beta)}$ *and* $\frac{\log b}{2} \frac{\Gamma(1+\gamma)\cdot\Gamma(1+\delta)}{\Gamma(1+\alpha)\cdot\Gamma(1+\beta)}$.

This result in a somewhat less general form has been proved from first principles by B. C. Das [1].

MADRAS CHRISTIAN COLLEGE.
TAMBARAM (MADRAS).
S. INDIA.

References

[1] Das, B. C., Effect of derangement of terms on some semi-convergent series, *Math. Student*, 5 (1937), 12–18.

[2] Littlewood, J. E., Note on the convergence of series of positive terms, *Messenger of Maths.*, 39 (1910), 191–192.

[3] Pringsheim, A., Allgemeine Theorie der Divergenz und Convergenz von Reihen mit Positiven Gliedern, *Math. Annalen*, 35 (1890), 297–394.

[4] Pringsheim, A., Zur Theorie der Dirichlet'schen Reihen, *Math. Annalen*, 37 (1890), 38–60.

[5] Pringsheim, A., Ueber die Werthveränderungen bedingt convergenter Reihen und Producte, *Math. Annalen*, 22 (1883), 455–503.

[6] Rajagopal, C. T., Convergence theorems for series of positive terms, *Journ. Indian Math. Soc.* (New Series), 3 (1938), 118–125.

[7] Rajagopal, C. T., On certain theorems of Pringsheim, *Tôhoku Math. Journ.*, 43 (1937), 122–126.

[8] Rajagopal, C. T., Some theorems connected with Maclaurin's integral test, *Math. Gazette*, 23 (1939), 456–461.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON MEAN ONE-VALENT FUNCTIONS

BY D. C. SPENCER[1]

(Received October 20, 1939)

1. In several recent papers[2] I have shown that many of the properties of **p**-valent functions are possessed by the wider family of functions which are **p**-valent on the average. The present paper is a study of the finer structure of functions one-valent on the average in which it is shown that such functions possess some of the more striking properties of schlicht functions. The method of the paper also yields two results of a more general nature which are stated in §'s (7) and (8) below (Theorems 3 and 4). The main results have been stated without proof elsewhere.[3]

2. For completeness we begin by restating the definition of mean **p**-valency. We suppose that $f(z)$ is regular in $|z| < 1$, and that W is the Riemann domain which is the transform of $|z| < 1$ by $f(z)$. Let $W(R)$ be the area (multiply covered regions being counted multiply) of that portion of W which lies in the circle $|w| \leqq R$. Then if

$$W(R) \leqq \mathbf{p}\pi R^2 \tag{2.1}$$

for all $R > 0$, we say that $f(z)$ is mean **p**-valent.

It is convenient to express the inequality (2.1) analytically. Let $n(r, w)$ be the number of times (necessarily bounded by some number depending upon r) that $f(z)$ takes the value w in $|z| < r$. We define

$$p(r, R) = \frac{1}{2\pi}\int_{-\pi}^{\pi} n(r, Re^{i\psi})\, d\psi; \tag{2.2}$$

$$n(w) = \lim_{r\to 1} n(r, w); \qquad p(R) = p(1, R) = \lim_{r\to 1} p(r, R). \tag{2.3}$$

Then

$$\begin{aligned} W(R) &= \lim_{r\to 1}\int_0^R \int_{-\pi}^{\pi} n(r, Re^{i\psi}) R\, dR\, d\psi \\ &= \int_0^R \left(\lim_{r\to 1}\frac{1}{2\pi}\int_{-\pi}^{\pi} n(r, Re^{i\psi})\, d\psi\right) d(\pi R^2) = \int_0^R p(R)\, d(\pi R^2), \end{aligned} \tag{2.4}$$

[1] I am indebted to Mr. K. J. Arnold for making the drawing which is reproduced in Fig. 4 below, on the original of which 679 values of the function $K(z) = 4z/(1+z)^2$ were plotted.

[2] For example, references (2) and (3) at the end of the paper.

[3] Spencer (5).

and so (2.1) may be written in the form

$$\int_0^R p(R)\, d(\pi R^2) \leqq \mathfrak{p}\pi R^2 \qquad (R > 0). \tag{2.5}$$

If we omit the averaging at (2.2), and suppose that

$$\int_0^R n(Re^{i\psi})\, d(\pi R^2) \leqq \mathfrak{p}\pi R^2 \tag{2.6}$$

for all $R > 0$, uniformly in ψ, then we say that f is strongly mean $\mathfrak{p}$-valent.[4] Fig. 1 represents a domain W which is strongly mean 1-valent (*a fortiori* mean 1-valent), but which is not schlicht.

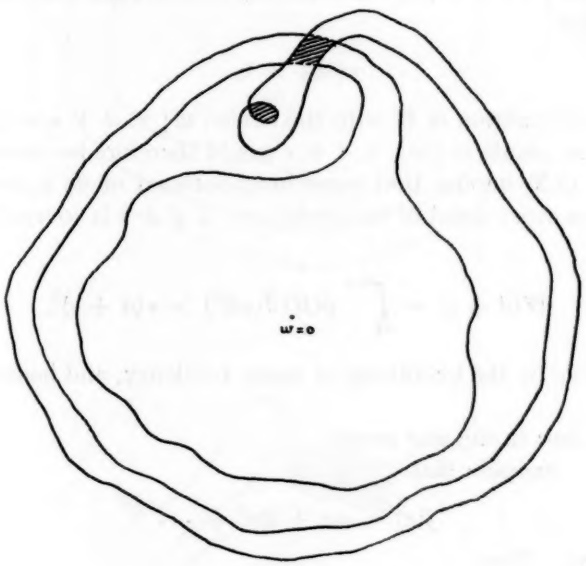


FIG. 1

3. It will be convenient to sum up here for future reference certain trivial deductions from mean 1-valency.

LEMMA 1. *Suppose that*

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is mean 1*-valent, and let*

$$d = \operatorname*{Inf}_{|z|<1} \left| \frac{f(z)}{z} \right|.$$

Then $d > 0$, *and* $n(w) = 1$ *if* $0 \leqq |w| < d$. *Furthermore, there exists a real number* β *such that* $n(de^{i\beta}) = 0$.

[4] Strong mean p-valency is therefore roughly midway in strength between p-valency and mean p-valency.

In fact, since $f(0) = 0$, there exists a circle of radius $\delta > 0$, and center $w = 0$, every point P of which is covered at least once by W. Since, by the definition of mean 1-valency, it is covered on the average at most once, it follows that the circle is covered exactly once by W. Hence d, the distance of the boundary of W from $w = 0$, satisfies $d \geqq \delta > 0$.

Suppose now that

$$n(de^{i\beta}) > 0 \tag{3.1}$$

for every β. To every β, $-\pi \leqq \beta < \pi$, there then corresponds a circle of radius $\rho = \rho(\beta)$ and center $de^{i\beta}$ every point of which is covered at least once, and it follows from the Heine-Borel theorem that there is a number $\epsilon > 0$ such that $n(w) \geqq 1$ for $|w| < d + \epsilon$. But there is then at least one point w_0, $|w_0| < d + \epsilon$, for which

$$n(w_0) > 1. \tag{3.2}$$

Otherwise the intersection of W with the circle $|w| < d + \epsilon$ would be schlicht, and its boundary points in $|w| < d + \epsilon$ would therefore be uncovered, contradicting (3.1). (3.2) implies that some neighborhood of w_0 is covered at least twice, and, since every point of the circle $|w| < d + \epsilon$ is covered at least once, we see that

$$W(d + \epsilon) = \int_0^{d+\epsilon} p(R)\, d(\pi R^2) > \pi(d + \epsilon)^2.$$

This is impossible by the hypothesis of mean 1-valency, and hence (3.1) is false.

4. We come now to the first result:

THEOREM 1. *Suppose that*

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is mean 1-valent. Then

$$|a_2| \leqq 2|a_1|, \tag{4.1}$$

and equality occurs if, and only if, $f(z) = a_1 z/(1 + \gamma z)^2$, $|\gamma| = 1$.

We write $z = re^{i\varphi}$, $w = f(z) = Re^{i\Phi}$. Then Theorem 1 follows quite easily from the following lemma, which is a generalization of the Bierberbach-Faber *Flächensatz*.

LEMMA 2. *If* $f(z)$ *is mean 1-valent and if* $\alpha > 0$, *then*

$$\int_{|z|=r} R^{-\alpha}\, d\Phi \geqq 0 \qquad (r < 1). \tag{4.2}$$

Lemma 2 when f is schlicht has been given by Prawitz,[5] who has used it to prove (4.1) for schlicht functions. The deduction of (4.1) from Lemma 2 is the same in our case, and I omit the calculations.

[5] See reference (2) at the end of the paper.

It remains, then, to prove Lemma 2. We suppose that $0 < r < 1$. Let $W(r)$ be the transform of $|z| < r$ by f, $B(r)$ its boundary (so that $B(r)$ is the transform of $|z| = r$); and let $d(r)$ (see Fig. 2)[6] be the distance of $B(r)$ from $w = 0$. Since the function $f(rz)$ (with picture $W(r)$) is *a fortiori* mean 1-valent

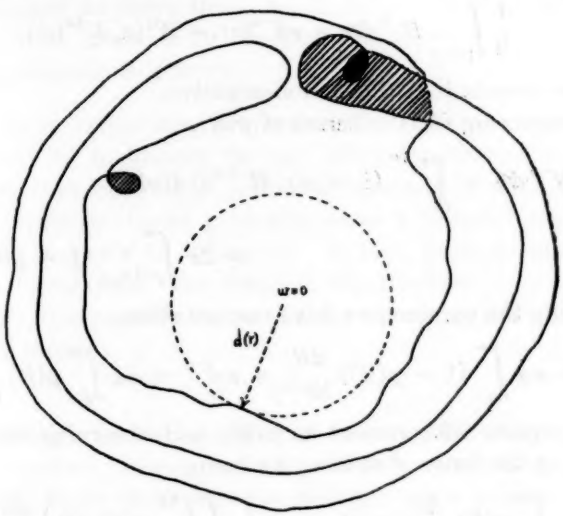


FIG. 2

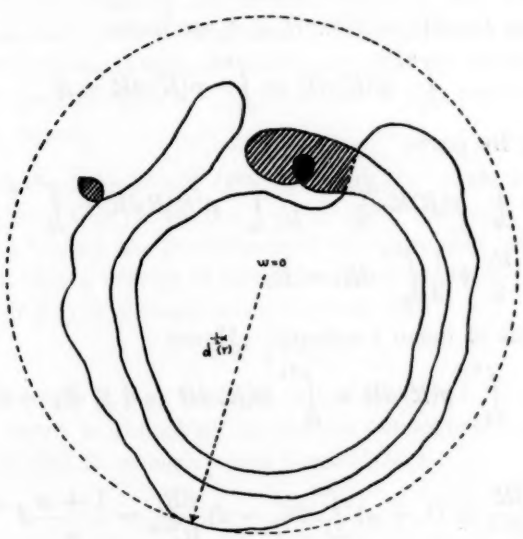


FIG. 3

and since $f(0) = 0$, $d(r) > 0$ by Lemma 1. Let $P = Re^{i\Phi}$ be a point of $W(r)$, and let $W'(r)$ (see Fig. 3) be the domain composed of the points $P' = R^{-\frac{1}{2}\alpha}e^{i\Phi}$.

[6] Incidentally, Fig. 2 represents a W which is mean 1-valent without being strongly mean 1-valent.

Then $W'(r)$ contains the point $w = \infty$, but is bounded ("internally") by a closed (but not necessarily simple) curve $B'(r)$ lying in the circle $|w| \leqq [d(r)]^{-\frac{1}{2}\alpha}$. Let $W'(r, R)$ be the area (multiply covered areas being counted multiply) of that portion of $W'(r)$ which lies in the circle $|w| < R$. Then

$$\frac{1}{2}\int_{|z|=r} R^{-\alpha}\, d\Phi = \pi d^{-\alpha}(r) - W'(r, d^{-\frac{1}{2}\alpha}(r)), \tag{4.3}$$

and the lemma asserts that this is non-negative.

We have, expressing (4.3) in terms of $p(R)$,

$$\frac{1}{2}\int_{|z|=r} R^{-\alpha}\, d\Phi = \int_0^{d^{-\frac{1}{2}\alpha}} (1 - p(r, R^{-2/\alpha}))\, d(\pi R^2)$$

$$= 2\pi \int_{d^{-\frac{1}{2}\alpha(r)}}^{\infty} (1 - p(r, R^{2/\alpha}))\frac{dR}{R^3},$$

and so, dropping the parameter r from our notation,

$$= \pi\alpha \int_d^{\infty} (1 - p(R))\frac{dR}{R^{1+\alpha}} = \pi d^{-\alpha} - \pi\alpha \int_d^{\infty} p(R)\frac{dR}{R^{1+\alpha}}. \tag{4.4}$$

Integrating by parts with respect to $p(R)$, and observing that the integrated part vanishes at the limits d and ∞, we have

$$\int_d^{\infty} p(R)\frac{dR}{R^{1+\alpha}} = (1+\alpha)\int_d^{\infty}\left(\int_d^{R_1} p(R)\, dR\right)\frac{dR_1}{R_1^{2+\alpha}}.$$

Since, by Lemma 1, $p(R) = 1$ for $R < d$, we have

$$\int_d^{R_1} p(R)\, dR = \int_0^{R_1} p(R)\, dR - d.$$

But, integrating by parts,

$$\int_0^{R_1} p(R)\, dR = \int_0^{R_1} p(R)R\cdot\frac{dR}{R} = \frac{1}{R_1}\int_0^{R_1} p(R)R\, dR + \int_0^{R_1}\left(\int_0^{R} p(\underline{R})\underline{R}\, d\underline{R}\right)\frac{dR}{R^2}$$

$$\leqq \frac{R_1}{2} + \frac{1}{2}\int_0^{R_1} dR_1 = R_1$$

by the hypothesis of mean 1-valency. Hence

$$\int_d^{R_1} p(R)\, dR = \int_0^{R_1} p(R)\, dR - d \leqq R_1 - d,$$

and so

$$\int_d^{\infty} p(R)\frac{dR}{R^{1+\alpha}} \leqq (1+\alpha)\int_d^{\infty}(R_1 - d)\frac{dR_1}{R_1^{2+\alpha}} = \frac{1+\alpha}{\alpha}d^{-\alpha} - d^{-\alpha} = \frac{1}{\alpha}d^{-\alpha}. \tag{4.5}$$

Substituting from (4.5) into (4.4), we obtain

$$\frac{1}{2}\int_{|z|=r} R^{-\alpha}\, d\Phi \geqq \pi d^{-\alpha}(r) - \pi d^{-\alpha}(r) = 0,$$

and this proves Lemma 2.

By a suitable choice of α in Lemma 2 it can be shown (as Prawitz has pointed out in the case of schlicht functions) that

$$|a_3| \leqq 3.05\,|a_1|.$$

I have not attempted to prove that $|a_3| \leqq 3\,|a_1|$ (with equality only for the Koebe function), but it seems to me that this result must certainly be true (though possibly difficult to prove).

5. The two most important properties of 1-valency (more generally of p-valency) are: (i) an invariance to one-valent transformations; and (ii) the metrical property that large areas cannot accumulate in small neighborhoods.[7] Of these mean 1-valency (more generally mean p-valency) possesses only the second; the group property (i) is lacking. In fact, mean p-valency is not even an invariant to translations, the simplest transformations of all. When f is 1-valent (schlicht), it is possible by use of the property (i) to deduce immediately from Theorem 1 that

$$d \geqq \tfrac{1}{4}\,|a_1|, \tag{5.1}$$

where d is the distance of the boundary of the map W from $w = 0$; but when f is only mean 1-valent the argument breaks down completely. The result is that I am able to prove (5.1) only for strongly mean 1-valent functions, and the precise inequality for mean 1-valent functions remains undetermined. *However, under mean 1-valency the weaker result is true that* $d \geqq |a_1|/A$, *where* A *is an absolute constant less than* 7; and more detailed computations indicate (but in their present form do not prove) that the result is still true when A is only slightly greater than 4. A discussion of this problem is to be found at the end of the paper.

6. We now take up the result (5.1) for strongly mean 1-valent functions. Afterwards we state and prove the two general theorems mentioned in §1 (Theorems 3 and 4 below), the hypotheses of which are essentially the conclusion of Lemma 1 (here only a vestige of mean 1-valency appears).

THEOREM 2. *If* $f(z)$ *is strongly mean 1-valent, then*

$$d \geqq \tfrac{1}{4}\,|a_1|$$

with equality if, and only if, $f(z) = a_1 z/(1 + \gamma z)^2$, $|\gamma| = 1$.

The following result is almost an immediate consequence of Theorem 2:

COROLLARY. *If* $f(z)$ *is strongly mean 1-valent, then*

$$|f(z)| \geqq \frac{|a_1| \cdot |z|}{(1 + |z|)^2},$$

and there is equality if, and only if,

$$f(z) = a_1 z/(1 + \gamma z)^2, \qquad |\gamma| = 1.$$

[7] It is the distribution of the area of the map with respect to $w = 0$ which determines the magnitude of the mean values of $|f|$ and $|f'|$ (and so of the upper bounds of the coefficients).

In fact, more generally, the inequality

$$|f(z)| \geqq \frac{4}{A} \cdot \frac{|a_1| \cdot |z|}{(1+|z|)^2}$$

is a consequence of $d \geqq |a_1|/A$ for *any* f. I omit a proof of this statement since the proof is substantially the same as in the well-known case when f is schlicht (see R. Nevanlinna, *Eindentige analytische Funktionen*, Berlin, 1936).

In the sequel we shall make repeated use of the following lemma. Suppose that

$$g(z) = b_1 z + b_2 z^2 + \cdots$$

is regular in $|z| < 1$. Let G be the map of the unit circle by g, and let G^* be the "star" of G with respect to the point $g(0)$.[8] We write $Z^* = g^{-1}(G^*)$, and denote by $p_1^*(R)$, $p_2^*(R)$ the valency functions of G^* and Z^* respectively.[9]

LEMMA 3.

$$\lg |b_1| < \int_0^\infty \lg R \, d[-p_1^*(R)] + \int_0^\infty \lg R \, d[-p_2^*(R)] \tag{6.1}$$

unless $g = b_1 z$.

(6.1), in different notation, is an inequality of Bermant.[10]

The lemma remains true if g is regular only in Z^*, but is false in general if this condition is not satisfied.

To prove Theorem 2 we suppose that $d = 1$, and prove that then $|a_1| \leqq 4$, a statement which is plainly equivalent to the theorem. By lemma 1 there exists a β such that $f \neq e^{i\beta}$ in $|z| < 1$. We write

$$K_\beta(z) = 4e^{i\beta} z/(1+z)^2.$$

This function (Koebe's function) maps the circle $|z| < 1$ on the plane slit along a straight line from $e^{i\beta}$ to ∞, and the inverse function $K_\beta^{-1}(w)$ has branch points of order 2 at $w = e^{i\beta}$ and at ∞. Since $f \neq e^{i\beta}, \infty$, however, either branch of K_β^{-1} is regular and one-valued on W, the map of $|z| < 1$ by f. To fix ideas we suppose that $K_\beta^{-1}(w)$ is the continuation over W of the branch which vanishes at $w = 0$,[11] and write

$$g(z) = K_\beta^{-1}\{f(z)\} = b_1 z + \cdots,$$

where $|b_1| = |a_1|/4$. By Lemma 3 we have

$$\lg \frac{|a_1|}{4} \leqq \int_0^\infty \lg R \, d[-p_1^*(R)] + \int_0^\infty \lg R \, d[-p_2^*(R)], \tag{6.2}$$

---

[8] More precisely, G^* is the subdomain of G any point of which can be reached by starting at $g(0)$ and travelling outward along some radius vector. G^* does *not* depend on the particular way in which the sheets of G are supposed stuck together.

[9] $p^*(R)$, for example, is therefore the average valency of G^* on the circumference $|g| = R$. Both G^* and Z^* are schlicht, and so p^* and p_2^* are bounded above by unity.

[10] For a proof of (6.1) in the above form see Spencer (5).

[11] We might equally well take the other branch (which is infinite at the origin).

and it is sufficient to show that

$$\int_0^\infty \lg R\, d[-p_1^*(R)] \leqq 0 \tag{6.3}$$

$$\int_0^\infty \lg R\, d[-p_2^*(R)] \leqq 0. \tag{6.4}$$

(6.4) is immediate. For Z^* is a schlicht domain which contains the point $z = 0$, and lies inside $|z| < 1$. Therefore by a known inequality[12] we have

$$\exp\left\{\int_0^\infty \lg R^2\, d[-p_2^*(R)]\right\} \leqq \int_0^\infty R^2\, d[-p_2^*(R)] = \frac{1}{\pi}\int_0^\infty p_2^*(R)\, d(\pi R^2) \leqq 1,$$

and this is equivalent to (6.4).

As for (6.3), we prove the following stronger result that

$$\int_0^\infty \lg R\, d[-p_1(R)] \leqq 0, \tag{6.5}$$

where $p_1(R)$ is the valency function of G.

Let $\mathfrak{E}(R_0)$ be the set of points in $|z| < 1$ at which $|g| = |K_\beta^{-1}\{f\}| > R_0$, and write

$$A(\beta, R_0) = \iint_{\mathfrak{E}(R_0)} \left|\frac{d}{dz}\lg g\right|^2 r\, dr\, d\varphi = \iint_{\mathfrak{E}(R_0)} \frac{|f'(re^{i\varphi})|^2}{|f(re^{i\varphi}) - e^{i\beta}|\cdot|f(re^{i\varphi})|^2}\, r\, dr\, d\varphi.$$

We suppose that R_0 is small enough to satisfy any condition imposed by the argument. Then $p_1(R) = 1$ for $R \leqq R_0$ by Lemma 1, and since $p_1(\infty) = 0$ we have, integrating by parts,

$$\begin{aligned} 2\pi\int_0^\infty \lg R\, d[-p_1(R)] &= 2\pi\int_{R_0}^\infty \lg R\, d[-p_1(R)] \\ &= \int_{R_0}^\infty p_1(R)\, d(2\pi \lg R) - 2\pi \lg\frac{1}{R_0} = A(\beta, R_0) - 2\pi\lg\frac{1}{R_0}. \end{aligned} \tag{6.6}$$

Next,

$$\begin{aligned} &\left|\iint_{\mathfrak{E}(R_0)} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi - \int_0^1\int_{-\pi}^{\pi}_{|f|>4R_0} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi\right| \\ &\qquad \leqq \frac{C}{R_0^2}\int_0^1\int_{-\pi}^{\pi}_{4R_0 - CR_0^2 < |f| < 4R_0 + CR_0^2} |f'|^2\, r\, dr\, d\varphi = \frac{C}{R_0^2}\int_{4R_0 - CR_0^2}^{4R_0 + CR_0^2} p(R)\, d(\pi R^2) \\ &\qquad = \frac{C}{R_0^2}\int_{4R_0 - CR_0^2}^{4R_0 + CR_0^2} d(\pi R^2) = O(R_0), \end{aligned}$$

by Lemma 1 and the hypothesis that $d = 1$; and hence

$$A(\beta, R_0) = \int_0^1\int_{-\pi}^{\pi}_{|f|>4R_0} \frac{|f'|^2}{|f - e^{i\beta}|\cdot|f|^2}\, r\, dr\, d\varphi + O(R_0). \tag{6.7}$$

[12] See Spencer (5).

But

$$\int_0^1 \int_{-\pi}^{\pi} \frac{|f'|^2}{|f - e^{i\beta}| \cdot |f|^2} r\,dr\,d\varphi = \int_{4R_0}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i\Phi})}{|e^{i\beta} - Re^{i\Phi}| \cdot R^2} R\,dR\,d\Phi$$

$$= \int_{\substack{4R_0 \\ |f|>4R_0}}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i\Phi})\,dR\,d\Phi}{R\sqrt{1 - 2R\cos(\Phi - \beta) + R^2}} \tag{6.8}$$

$$= \int_{4R_0}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i(\Phi+\beta)})}{R\sqrt{1 - 2R\cos\Phi + R^2}}\,dR\,d\Phi.$$

Substituting from (6.8) into (6.7) we obtain the formula (valid for all sufficiently small R_0):

$$A(\beta, R_0) = \int_{4R_0}^{\infty} \int_{-\pi}^{\pi} \frac{n(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R\cos\Phi + R^2}}\,dR\,d\Phi + O(R_0). \tag{6.9}$$

Integrating by parts we have

$$\int_{4R_0}^{\infty} \frac{n(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R\cos\Phi + R^2}}\,dR = \left[\frac{1}{\sqrt{1 - 2R\cos\Phi + R^2}} \int_{4R_0}^{R} \frac{n}{\underline{R}}\,d\underline{R}\right]_{4R_0}^{\infty}$$

$$- \int_{4R_0}^{\infty} \left(\int_{4R_0}^{R} \frac{n}{\underline{R}}\,d\underline{R}\right) \frac{d}{dR}\left\{\frac{1}{\sqrt{1 - 2R\cos\Phi + R^2}}\right\} dR$$

$$= -\int_{4R_0}^{\infty} \left(\int_{4R_0}^{R} \frac{n}{\underline{R}}\,d\underline{R}\right) \frac{d}{dR}\left\{\frac{1}{\sqrt{1 - 2R\cos\Phi + R^2}}\right\} dR \tag{6.10}$$

$$= \int_{4R_0}^{1} \left(\int_{4R_0}^{R} \frac{1}{\underline{R}}\,d\underline{R}\right) \frac{R - \cos\Phi}{(1 - 2R\cos\Phi + R^2)^{3/2}}\,dR$$

$$+ \int_{1}^{\infty} \left(\int_{4R_0}^{R} \frac{n}{\underline{R}}\,d\underline{R}\right) \frac{R - \cos\Phi}{(1 - 2R\cos\Phi + R^2)^{3/2}}\,dR,$$

since (by hypothesis) $d = 1$ and so, by Lemma 1, $n = 1$ for $0 \leqq R < 1$. We write

$$I(R_0, \Phi) = \int_{4R_0}^{1} \left(\int_{4R_0}^{R} \frac{1}{\underline{R}}\,d\underline{R}\right) \frac{R - \cos\Phi}{(1 - 2R\cos\Phi + R^2)^{3/2}}\,dR. \tag{6.11}$$

Then $I(R_0, \Phi)$ depends only on R_0 and Φ, but *not* on f. Next, writing

$$N(R, \Phi + \beta) = \int_{4R_0}^{R} \frac{n(\underline{R}e^{i(\Phi+\beta)})}{\underline{R}}\,d\underline{R}, \tag{6.12}$$

we have

$$N(R, \Phi + \beta) = \left[\frac{1}{R}\int_{4R_0}^{R} n\,d\underline{R}\right]_{4R_0}^{R} + \int_{4R_0}^{R} \left(\int_{4R_0}^{R} n\,d\underline{R}\right) \frac{dR_1}{R_1^2}$$

$$\leqq \frac{1}{R}(R - 4R_0) + \int_{4R_0}^{R} (R_1 - 4R_0) \frac{dR_1}{R_1^2} = \lg \frac{R}{4R_0}$$

by the hypothesis of strong mean 1-valency and the fact that $n = 1$ if $R < 1$ (since $d = 1$). Hence

$$N(R, \Phi + \beta) \leqq \int_{4R_0}^{R} \frac{dR}{R} = \lg \frac{R}{4R_0}, \tag{6.13}$$

uniformly in Φ. Since

$$\frac{R - \cos \Phi}{(R^2 - 2R \cos \Phi + 1)^{3/2}} > 0$$

if $R > 1$, we have, by (6.10), (6.11), (6.12), and (6.13), that

$$\int_{4R_0}^{\infty} \frac{n(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}}\, dR \leqq \int_{4R_0}^{\infty} \frac{1}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}}\, dR. \tag{6.14}$$

Integrating both sides of (6.14) with respect to Φ from $-\pi$ to π, we obtain the inequality

$$\begin{aligned} A(\beta, R_0) &= \int_{-\pi}^{\pi} \int_{4R_0}^{\infty} \frac{n(Re^{i(\Phi+\beta)})}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}}\, dR\, d\Phi + O(R_0) \\ &\leqq \int_{-\pi}^{\pi} \int_{4R_0}^{\infty} \frac{1}{R \cdot \sqrt{1 - 2R \cos \Phi + R^2}}\, dR\, d\Phi + O(R_0) \qquad (6.15) \\ &= \int_{-\pi}^{\pi} \int_{R_0}^{1} \left|\frac{1}{z}\right|^2 r\, dr\, d\varphi + 0(R_0) = 2\pi \lg \frac{1}{R_0} + O(R_0), \end{aligned}$$

since when n is replaced by unity, we may suppose that $f = K_\beta$, and so $g = z$.

Finally, substituting from (6.15) into (6.6), we have that

$$\int_0^{\infty} \lg R\, d[-p_1(R)] \leqq O(R_0).$$

Letting $R_0 \to 0$, we obtain (6.5), and this completes the proof of Theorem 2 apart from the statement concerning equality. For this we note that equality implies by Lemma 3 that $g = z$, therefore that $f = K_\beta$.

7. Suppose that $f(0) = 0$. By Lemma 1 mean 1-valency then implies: (i) that $n(0) = 1$; and (ii) that $n(de^{i\beta}) = 0$ for at least one β. We now take for hypotheses only (i) and (ii), and obtain the following theorem:

THEOREM 3. *Suppose that*

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is regular for $|z| < 1$, *that* $n(0) = 1$, *and that*

$$d = \operatorname*{Inf}_{|z|<1} \left|\frac{f(z)}{z}\right|.$$

Then if there is a β *for which* $n(de^{i\beta}) = 0$, *we have*

$$d > \frac{1}{13} |a_1|. \tag{7.1}$$

The constant 13 is not the best possible. As a matter of fact, a better (and in a sense best possible) constant is obtainable by a "subordination" argument (Lindelöf's principle). But the theorem is only mildly interesting on its own account, and is included here mainly in order to deduce Theorem 4 below. The hypothesis that $n(de^{i\beta}) = 0$ for some β is necessary. This is seen by considering the function

$$f(z) = z - z^2,$$

for which $n(de^{i\beta}) = n(0) = 1$.

Since by hypothesis $n(0) = 1$, $n(de^{i\beta}) = 0$, we have $d > 0$. We suppose that $d = 1$, and prove that $|a_1| < 13$. Then if β is a number for which $n(e^{i\beta}) = 0$, the function[13] $g = K_\beta^{-1}\{f\}$ is regular for $|z| < 1$. By Lemma 3 we have, therefore,

$$\lg \frac{|a_1|}{4} \leqq \int_0^\infty \lg R \, d[-p_1^*(R)] + \int_0^\infty \lg R \, d[-p_2^*(R)] \leqq \int_0^\infty \lg R \, d[-p_1^*(R)] = J^*, \text{ say,} \tag{7.2}$$

where $p_1^*(R)$ is the valency function of G^*. We find an upper bound for J^*.

Now every point of the circle $|w| < 1$ is covered exactly once. For $n(0) = 1$, and if some point w_0, $|w_0| < 1$, were covered n-times, where $n \neq 1$, then the projection of a boundary point would fall on the straight line connecting 0 with w_0; and this is impossible since $d = 1$.

We define

$$t_\beta(\varphi) = \begin{cases} K_\beta\{\mathfrak{E}[\arg z = \varphi, 0 \leqq |z| < \infty]\}, & \varphi \neq 0, \pi \\ K_\beta\{\mathfrak{E}[\arg z = \varphi, 0 \leqq |z| < 1]\}, & \varphi = 0, \pi \end{cases} \tag{7.3}$$

where $\mathfrak{E}[\cdots]$ is the set of points defined by the conditions stated in the bracket. Positions lying outside $|w| < 1$ of the lines t_0 which cut $|w| = 1$ at intervals of 5° from 0° to 180° are shown in Fig. 4. We denote by W_β^* the subdomain of W any point P of which can be reached by starting at $w = 0$ and travelling outward along some line $t_\beta(\varphi)$, $\varphi = \varphi(P)$, always in the sense of increasing $|z|$. Then $G^* = K_\beta^{-1}\{W_\beta^*\}$. Also if we let $n_\beta^*(w)$ be the number of times W_β^* covers the point w, then

$$n_\beta^*(w) \leqq 2 \tag{7.4}$$

for all w.

Next, if R_0 is sufficiently small, we have[14]

$$2\pi J^* = \int_{4R_0}^\infty \int_{-\pi}^{\pi} \frac{n_\beta^*(Re^{i(\Phi+\beta)})}{R\cdot\sqrt{1 - 2R\cos\Phi + R^2}} \, dR \, d\Phi - 2\pi \lg \frac{1}{R_0} + O(R_0). \tag{7.5}$$

That is to say, $2\pi J^*$ is the difference between the logarithmic area of G^* and

[13] The branch of K_β^{-1} which vanishes at $w = 0$ is again chosen.

[14] By an argument analogous to the one given in §6.

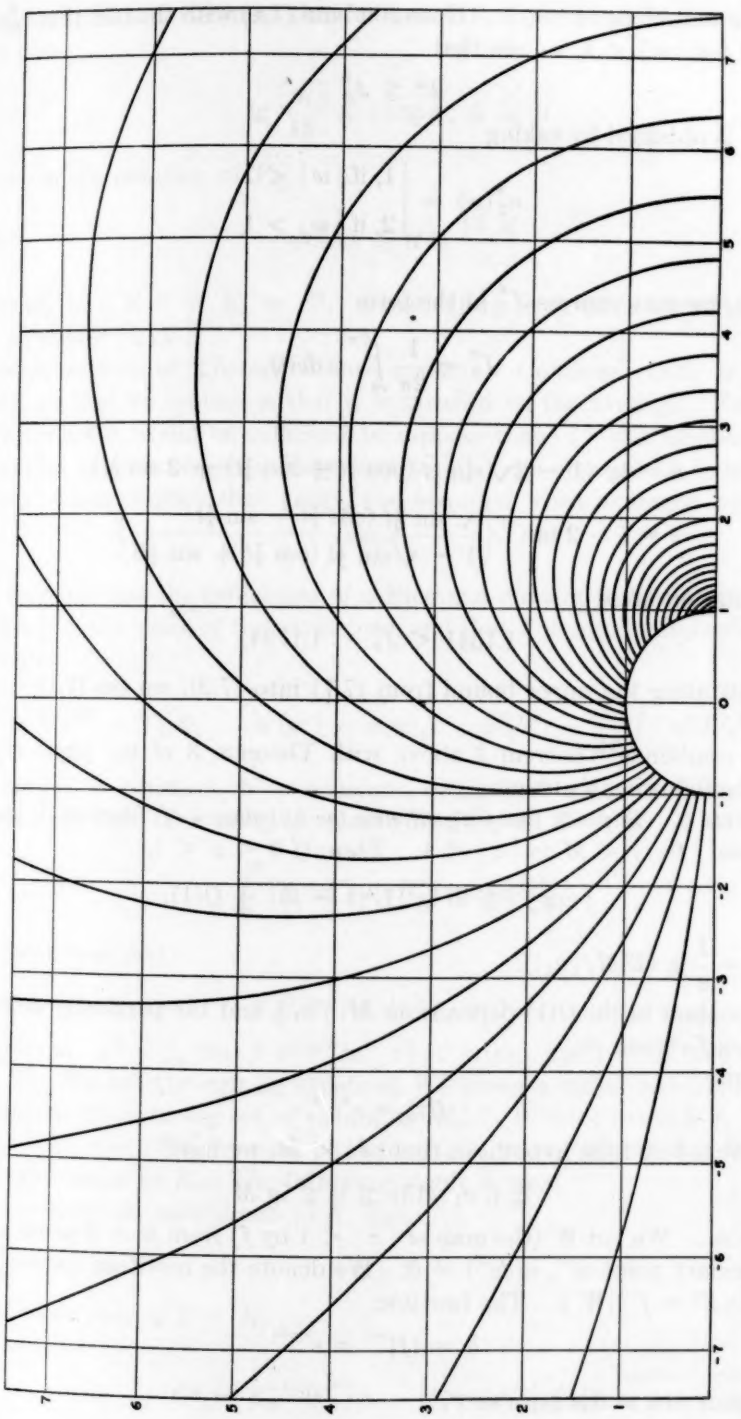


FIG. 4

that of the unit circle $|g| < 1$. If we combine (7.4) with the fact that $n_\beta^*(w) = n(w) = 1$ for $|w| < 1$, we see that

$$J^* \leqq J_0^*, \tag{7.6}$$

where J_0^* is obtained by taking

$$n_\beta^*(w) = \begin{cases} 1, \text{ if } |w| < 1 \\ 2, \text{ if } |w| > 1 \end{cases}$$

in (7.5)

Finally, we may express J_0^* in the form

$$J_0^* = \frac{1}{2\pi} \int_0^{2\pi} u \, dv(t),$$

where

$$u = \lg \{1 - 2\sqrt{\sin \tfrac{1}{2}t} \, (\cos \tfrac{1}{4}t + \sin \tfrac{1}{4}t) + 2 \sin \tfrac{1}{2}t\}.$$

$$v = t - 2 \tan^{-1}\left\{\frac{\sqrt{\sin \tfrac{1}{2}t} \, (\cos \tfrac{1}{4}t - \sin \tfrac{1}{4}t}{1 - \sqrt{\sin \tfrac{1}{2}t} \, (\cos \tfrac{1}{4}t + \sin \tfrac{1}{4}t)}\right\}.$$

A computation gives

$$1.1541 < J_0^* < 1.1784, \tag{7.7}$$

and substituting the upper bound from (7.7) into (7.2), we get (7.1).

8. By combining Theorem 3 above with Theorem 3 of my paper (4)[15], we obtain the following theorem:

THEOREM 4. *Suppose that $f(z)$ satisfies the hypotheses of Theorem 3, and that, in addition, $|f(z)| < M$ in $|z| < 1$. Then, if $0 < \rho < 1$,*

$$|\arg f| \leqq B \lg \{1/(1 - \rho)\} + O(1), \tag{8.1}$$

where $B = \dfrac{1}{\pi} \lg (13M/|a_1|)$.

The constant in the $O(1)$ depends on M, $|a_1|$, and the particular determination of $\arg f$ chosen.

We write

$$f(z) = e^{u+iv}.$$

By Theorem 3 and the hypothesis that $|f| < M$, we have

$$\lg (|a_1|/13) \leqq u \leqq \lg M, \tag{8.2}$$

if $|v| > v_0$. We cut W (the map of $|z| < 1$ by f) from $w = 0$ along a radius to a boundary point $de^{i\beta}$, $n(de^{i\beta}) = 0$. We denote the resulting domain by W', and write $Z' = f^{-1}\{W'\}$. The function

$$h = \{f\}^{-i} = e^{v-iu}$$

[15] We shall refer to this paper as V_2.

is regular in Z', and we apply the method of V_2 to the function h. By (8.2) we have that

$$\lg \frac{|a_1|}{13} \leqq |\arg h| \leqq \lg M.$$

Hence, in the notation of V_2,

$$\Theta_\nu(R) \leqq R \lg \frac{13\,M}{|a_1|} \tag{8.3}$$

uniformly in ν, if $R > R_0 = e^{\nu_0}$. Substituting from (8.3) into Theorem 3 of V_2, we obtain (8.1).

Generalizations of Theorem 4 are possible. Condition (8.2) is much too strong; all that we require is that it is satisfied on the average. For example, in Theorem 4 it would be sufficient to suppose that $|f|$, *qua* function of arg f, is bounded when averaged with respect to arg f (as opposed to ordinary mean valency, which implies that $|\arg f|$ is bounded when averaged with respect to $|f|$).

9. Suppose that the projection of a Riemann domain W lies inside a schlicht domain D in the space of the variable w, and that T is regular and schlicht in D. We write

$$w' = R'e^{i\Phi'} = T(w), \qquad n'(w') = n(w), \qquad p'(R') = \frac{1}{2\pi}\int_{-\pi}^{\pi} n'(R'e^{i\Phi'})\,d\Phi'.$$

The lines $R' =$ constant, $\Phi' =$ constant, define in D a new coördinate system. We shall say that W is mean **p**-valent (T) if for all $R' > 0$,

$$\int_0^{R'} p'(R')\,d(\pi R'^2) \leqq p\pi R'^2.$$

Suppose now that

$$f(z) = a_1 z + a_2 z^2 + \cdots$$

is regular in $|z| < 1$, that $f \neq e^{i\beta}$ in $|z| < 1$ (i.e., $n(e^{i\beta}) = 0$), and define W_β^* as in §7. We cut through all sheets of W_β^* along a radial line from $e^{i\beta}$ to ∞, and denote the resulting set of points by $\mathfrak{W}_\beta^*$. If there exists a β, $f \neq e^{i\beta}$ in $|z| < 1$, for which $\mathfrak{W}_\beta^*$ is mean 1-valent (K_β^{-1}), we say briefly that W is metrically subordinate to K. The following result is true:

If f is metrically subordinate to K, then

$$|a_1| \leqq 4, \tag{9.1}$$

with equality only if $f = K_\beta$.

In fact, let

$$G^* = K_\beta^{-1}\{W_\beta^*\}; \qquad \mathfrak{G}^* = K_\beta^{-1}\{\mathfrak{W}_\beta^*\}.$$

Then if $p_1^*(R)$, $\mathfrak{p}_1^*(R)$ are the valency functions of G^*, $\mathfrak{G}^*$ respectively, we have

$$\mathfrak{p}_1^*(R) = \begin{cases} p_1^*(R) + p_1^*\left(\frac{1}{R}\right), & \text{if } 0 \leqq R < 1, \\ 0, & \text{if } R > 1. \end{cases} \tag{9.2}$$

Furthermore, by hypothesis,

$$\int_0^R \mathfrak{p}_1^*(R) d(R^2) \leqq \begin{cases} R^2, & \text{if } R < 1 \\ 1, & \text{if } R > 1. \end{cases} \tag{9.3}$$

In particular, therefore, by Lemma 1, $\mathfrak{p}_1^*(R) = 1$ if $0 \leqq R < R_0$, and so (integrating by parts)

$$\begin{aligned} \int_0^\infty \lg R^2 \, d[-\mathfrak{p}_1^*(R)] &= \lg R_0^2 + \int_{R_0}^\infty \mathfrak{p}_1^*(R) \frac{d(R^2)}{R^2} \\ &= \lg R_0^2 + 2\int_{R_0}^\infty \left\{\int_{R_0}^R \mathfrak{p}_1^*(R)\, d(R^2)\right\} \frac{dR}{R^3} \\ &= \lg R_0^2 + 2\int_{R_0}^1 \left\{\int_{R_0}^R \mathfrak{p}_1^*(R) d(R^2)\right\} \frac{dR}{R^3} + 2\int_1^\infty \left\{\int_{R_0}^R \mathfrak{p}_1^*(R)\, d(R^2)\right\} \frac{dR}{R^3} \\ &= \lg R_0^2 + I_1 + I_2, \qquad \text{say,} \end{aligned} \tag{9.4}$$

where by (9.3)

$$I_1 = 2\int_{R_0}^1 \left\{\int_0^R \mathfrak{p}_1^*(R)\, d(R^2) - R_0^2\right\} \frac{dR}{R^3} \leqq -\lg R_0^2 + R_0^2 - 1 \tag{9.5}$$

$$I_2 \leqq 2\int_1^\infty \{1 - R_0^2\} \frac{dR}{R^3} = 1 - R_0^2. \tag{9.6}$$

Using (9.2) and substituting from (9.5) and (9.6) into (9.4), we obtain that

$$\begin{aligned} J^* = \int_0^\infty \lg R \, d[-p_1^*(R)] &= \int_0^1 \lg R \, d[-p_1^*(R)] + \int_1^\infty \lg R \, d[-p_1^*(R)] \\ &= \int_0^1 \lg R \, d[-p_1^*(R)] + \int_0^1 \lg R \, d\left[-p_1^*\left(\frac{1}{R}\right)\right] = \tfrac{1}{2}\int_0^\infty \lg R^2 \, d[-\mathfrak{p}_1^*(R)] \leqq 0, \end{aligned}$$

and this is equivalent, by Lemma 3, to (9.1).

10. The question now arises as to whether the result of Theorem 2 is true for mean 1-valent functions. I am not able to answer this question, the difficulty being as follows.

Let F_1 be the family of mean 1-valent functions, and let

$$J_1 = \operatorname*{Max}_{f \in F_1} \int_0^\infty \lg R \, d[-p_1(R)], \qquad J_1^* = \operatorname*{Max}_{f \in F_1} \int_0^\infty \lg R \, d[-p_1^*(R)],$$

where $p_1(R)$, $p_1^*(R)$ are the valency functions of G and G^* respectively (so that $J_1^* = \operatorname{Max} J^*$). Then

$$0 < J_1^* < J_1 . \tag{10.1}$$

In fact, let W be the domain represented in Fig. 5, when $n(w) = 1$ for almost all w exterior to D_1 and D_2, and $n(w) = 2$ for $w \in D_1$, $n(w) = 0$ for $w \in D_2$. The solid curved lines are (we suppose) lines $t_0(\varphi)$. The boundary of W consists of the boundaries of D_1 and D_2, and slits. Let $D_1(R)$, $D_2(R)$ be the widths of D_1 and D_2 on $|w| = R$. We suppose: (i) that $D_1(R) = D_2(R)$ for all R; (ii) that

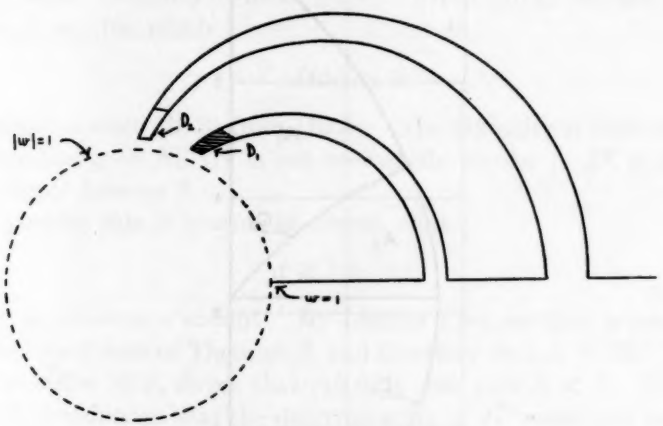


FIG. 5

all of W is swept out by lines $t_0(\varphi)$ as φ varies from 0 to 2π—that is to say, $W = W_0^*$. Both assumptions may plainly be fulfilled if D_1, D_2 have suitable shapes, and (i) implies that $p(R) = 1$ for all $R > 0$, so that W is mean 1-valent. Now write $w = Re^{i\Phi}$, $dW_w = R \, dR \, d\Phi$, and let dW_σ be the transform of dW_w by $K_\beta^{-1}(w)$. Then if $\mathfrak{R}$ is the distance from the element dW_w to the point $e^{i\beta}$, we have by (6.8) with $\beta = 0$ that

$$dW_\sigma = dW_w/\Lambda, \qquad \text{say,}$$

where $\Lambda = R^2\mathfrak{R}$. By sliding points $w = Re^{i\Phi}$ from one of the two sheets covering D_1 along the circumference $|w| = R$ into D_2, we can just fill the latter domain, and still have D_1 covered once. For the resulting picture $J^* = 0$. Since $\Lambda(Re^{i\Phi})$ is a *decreasing* function of Φ as we slide w along the circumference $|w| = R$ from D_1 into D_2, we see that the original J^* corresponding to Fig. 5 must be positive; and (since $J^* \leqq J_1^*$) this proves the first half of (10.1).

As for the second half, we may plainly suppose (by a trivial rotation of each function f) that $\beta = 0$. Then it is only necessary to point out that an argument similar to the above shows that J_1 is the area which results from the choice

$$n_1(Re^{i\Phi}) = \begin{cases} 1, & 0 \leqq R < 1 \\ 2, & R > 1, \quad -\frac{\pi}{2} < \Phi < \frac{\pi}{2} \\ 0, & R > 1, \quad \frac{\pi}{2} \leqq \Phi \leqq \frac{3\pi}{2} \end{cases}; \tag{10.2}$$

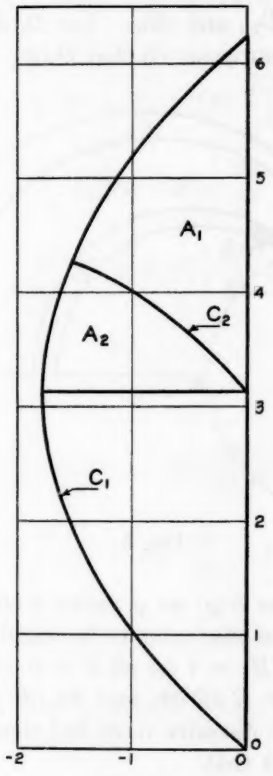


FIG. 6

In Fig. 6, C_1 is the transform of $|w| = 1$, C_2 the transform of any $w = 3\pi/2$ by $\lg K_0^{-1}(w)$, and $J_1 = \frac{1}{\pi}(A_1 - A_2)$. For any function n which satisfies mean 1-valency and differs from n_1 in a set of positive plane measure, the corresponding J is *less* than J_1. In particular, the function n^* corresponding to any W_0^*

plainly differs from n_1 in a set of positive plane measure;[16] and this proves the second half of (10.1).

Thus the method of this paper fails to yield the result

$$d \geqq \tfrac{1}{4} |\, a_1 \,|. \tag{10.3}$$

for the family F_1 of mean 1-valent functions.

It is perhaps worthwhile to point out that the area integral J (also J^*) is a function of the parameter β, and that $J(\beta)$ exist for all β whether $n(e^{i\beta}) = 0$ or not. This can be seen from the formula (6.9). Furthermore,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} J(\beta)\, d\beta \leqq 0 \tag{10.4}$$

if $f \in F_1$. In fact, the integration has the effect of replacing n by p in the formulas of §6, and the argument given there (with strong mean 1-valency replaced by mean 1-valency) yields (10.4). From (10.4) we see that we can choose $a\beta$, β_0 say, for which

$$J(\beta_0) \leqq 0. \tag{10.5}$$

But we cannot deduce (10.3) from (10.5). The difficulty is that for this choice of β the function $g = K_\beta^{-1}\{f\}$ is not necessarily regular in Z^*, and we cannot, therefore, apply Lemma 3.

On the positive side it is true, of course, that

$$d \geqq |\, a_1 \,|/A, \tag{10.6}$$

where A is an absolute constant. By Lemma 1 we see that a mean 1-valent f satisfies the hypotheses of Theorem 3, and therefore that $A < 13$. But a rather crude computation of J_1 shows that (10.6) is true with $A < 7$. Finally the inequality (10.1) indicates that the determination of J_1^* would still further reduce A. Although I have not attempted to determine J_1^*, I should like to conclude the paper by pointing out certain qualitative considerations which indicate (though of course they do not prove) that J_1^* (though positive) is small, and that (10.6) is therefore probably true with $A = 4$.

We suppose again (without loss of generality) that $\beta = 0$. The two lines $t_0(\varphi)$, $t_0(\varphi')$ which cut the circumference $|\, w \,| = 1$ at $\Phi - 1^0$ and at Φ respectively, $\Phi \geqq 1^0$, together with a portion of $|\, w \,| = 1$ and the real axis, bound a certain region of the w-plane which we shall denote by $\Delta(\Phi)$. We denote by $\alpha(\Phi)$ the area of the transform of $\Delta(\Phi)$ by $\lg K_0^{-1}(w)$. The function $\alpha(\Phi)$, labelled "area per degree", is plotted against Φ (in degrees) in Fig. 7. We note the following two facts: (i) $\alpha(\Phi)$ is an increasing function of Φ; and (ii) $\alpha(\Phi)$ varies

[16] If n^* is the "star" function derived from n_1, and $J^*(n_1^*)$ the corresponding "J^*", it is not difficult to show that $J^*(n^*) < 0$. If A_1^+ is the portion of A_1 which can be seen from $-\infty$, $J^*(n_1^*) = \dfrac{1}{\pi}\{2A_1^+ - (A_1 + A_2)\}$.

remarkably little, the variation being only a quarter (approximately) of its minimum.

Now the domain W_0^* is limited to a considerable extent by the shape of the lines t_0. Either W_0^* is schlicht, in which case $J^* \leqq 0$; or W_0^* covers some region D of the w-plane twice, in which case we suppose that D lies inside the region $\Delta(\Phi)$ and touches the two boundary lines $t_0(\varphi)$, $t_0(\varphi')$ of $\Delta(\Phi)$. Then every point of $\Delta(\Phi)$ (and its reflection in the real axis) is covered at least once. If now Φ is small, then the resulting contribution to J^* is below average by (i); if Φ is large, the contribution is relatively large, but mean 1-valency prevents too many

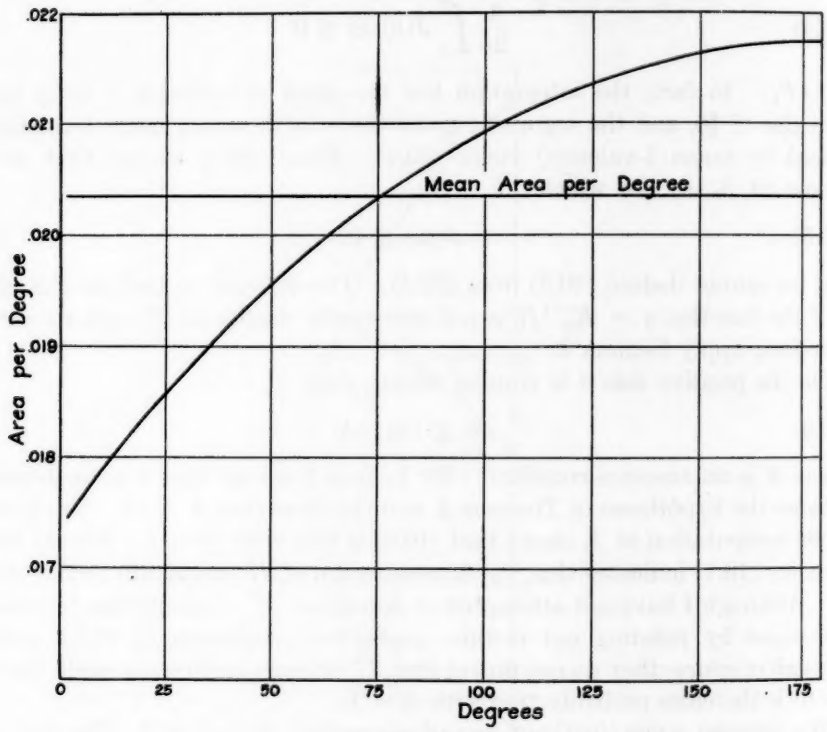


FIG. 7

D's from having this location, as can be seen from Fig. 4. In other words, mean 1-valency tends to prevent the accumulation of area in Δ's with large α's, and this is an indication that J_1^* is small.

As an example, let us consider the class C of functions f of F_1 for which W_0^* has the property that either no point of a region $\Delta(\Phi)$ is covered, or every point of it is covered twice. Then by mean 1-valency at most 180 regions $\Delta(\Phi)$ are occupied (since each must be filled twice). If we fill twice the regions $\Delta(\Phi)$, $\Phi = 90° + v$, $v = 1°, 2°, \cdots, 180°$, a precise computation shows that (10.6) is then true for this function with $A = 4.35$; and by property (i) above we see at once that (10.6) with $A = 4.35$ is true throughout the class C. For this result we have used mean

1-valency only on the circumference $|w| = 1$. If we use it everywhere, we find that (10.6) is true in C with $A = 4$.

11. In conclusion I should like to point out that results obtained for mean 1-valent functions may be immediately transferred to mean p-valent functions of the form

$$f(z) = \sum_{n=p}^{\infty} a_n z^n. \tag{11.1}$$

In fact, the function

$$f_1(z) = \{f(z)\}^{1/p}$$

is regular in $|z| < 1$ since the mean p-valency of f implies that $f = 0$ only at $z = 0$. Let $n_1(w)$ be the number of times f_1 takes the value w in $|z| < 1$, and let $p_1(R)$ be defined by (2.2) and (2.3) in terms of n_1. Then if $p(R)$ is the corresponding function for f, it is easy to see that

$$p_1(R) = \frac{1}{p} \cdot p(R^p), \tag{11.2}$$

and so

$$\begin{aligned} \int_0^R p_1(R)\, d(\pi R^2) &= \frac{1}{p} \int_0^R p(R^p)\, d(\pi R^2) \\ &= \frac{1}{p} \int_0^{R^p} p(R)\, d(\pi R^{2/p}) \leqq \frac{1}{p} \cdot p\pi R^2 = \pi R^2 \end{aligned} \tag{11.3}$$

by Lemma 1 of my paper (4). Hence f_1 is mean 1-valent, and so we have that $|a_{p+1}| \leqq 2p \cdot |a_p|$, $d = \operatorname{Inf}_{|z|<1} |f(z)/z^p| > |a_p|/7^p$, etc.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

(1) Bermant, A., "Remarque sur le lemme de Schwarz," *C. R. Acad. Sci. Paris*, 207 (1938), 31-33.

(2) Prawitz, H., "Über Mittlewerte analytischer Funktionen," *Arkiv för Mat.*, 20A (1927-28), 6, 1-12.

(3) Spencer, D. C., "On finitely mean valent functions," *Proc. London Math. Society* (to appear shortly).

(4) Spencer, D. C., "On finitely mean valent functions II," *Trans. Amer. Math. Society*, 48 (1940), 418-435.

(5) Spencer, D. C., "On an inequality of Gransky," *Proc. Nat. Acad. Sciences*, 26 (1940), 616-621.

(6) Spencer, D. C., "On distortion in analytic transformations," *Journal Math. and Physics, Mass. Inst. of Technology*, 20 (1941), 124-126.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

NECESSARY CONDITIONS IN THE THEORY OF INTERPOLATION IN THE COMPLEX DOMAIN

J. H. CURTISS

(Received May 18, 1940)

1. Introduction

Let C be a closed limited point set of the complex z-plane, and let A_C denote the class of functions analytic and single-valued on C. Let

$$M: \left\{ \begin{array}{l} \alpha_1^{(1)} \\ \alpha_1^{(2)}, \alpha_2^{(2)} \\ \vdots \\ \alpha_1^{(n)}, \alpha_2^{(n)}, \cdots, \alpha_n^{(n)} \\ \vdots \end{array} \right\}$$

be an infinite set of points of the finite plane which either lie on C or have no limit point except on C. (The points of M need not be all distinct.) A function $f(z) \in A_C$ will be analytic and single-valued at each of the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$, if n is sufficiently large, say $n > n(f)$. There exists a unique polynomial $L_n(z; f)$ of degree at most $n-1$, $n > n(f)$, found by interpolation to the function $f(z)$ in the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$.[1]

Certain conditions are easily proved to be sufficient for the convergence of the sequence $\{L_n(z; f)\}$ to $f(z)$ on C. The necessity of such conditions seems first to have been studied by Kalmár[2] in his thesis of 1926. He proved the following theorem:

THEOREM A. *Let C be a closed Jordan region and let the points M lie on the boundary B of C. Then a necessary condition that* $\lim_{n\to\infty} L_n(z; f) = f(z)$ *uniformly on C for each function $f(z) \in A_C$ is that the points M be uniformly-densely distributed on B.*

The sufficiency of the condition was proved earlier by Fejér.[3] A uniformly-dense distribution of the points M on the Jordan curve B is defined as follows: Let the analytic function $w = \varphi(z)$ map the complement of C (with respect to the extended plane) onto the region $|w| > 1$ so that the points at infinity

---

[1] See J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, New York, 1935, pp. 49–51.

[2] L. Kalmár, Über Interpolation. (In Hungarian.) Matematikai és Physikai Lapok, 1926, pp. 120–149.

[3] L. Fejér, Interpolation und konforme Abbildung, Göttinger Nachrichten, 1918, pp. 319–331.

correspond to each other. The mapping is bi-uniform and continuous for $|w| \geqq 1$. The points M are said to be uniformly-densely distributed on B if the points $\varphi(\alpha_k^{(n)})$, $k = 1, \cdots, n$, $n = 1, 2, \cdots$, are uniformly-densely distributed on the unit circle in the sense of Weyl;[4] that is, if for any arc of length l on the unit circle,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\varphi(\alpha_k^{(n)}) \subset l} 1 = \frac{l}{2\pi}.$$

Kalmár[5] also proved the following theorem in the case that the region K is simply connected:

THEOREM B. *Let the complement K of C be connected and regular in the sense that K possesses a Green's function $G(x, y)$ with pole at infinity. Let $\varphi(z) = e^{G+iH}$, where $H = H(x, y)$ is a harmonic conjugate of G, and let Δ be the transfinite diameter (or capacity) of C. Then a necessary condition that $\lim_{n\to\infty} L_n(z; f) = f(z)$ uniformly on C, for each function $f(z) \in A_C$, is that*

$$\lim_{n\to\infty} |\omega_n(z)|^{1/n} = \Delta\,|\varphi(z)|, \qquad \omega_n(z) = \prod_{k=1}^{n} (z - \alpha_k^{(n)}),$$

uniformly for z on any closed limited point set interior to K.

The sufficiency of the condition was also proved by Kalmár in the case that K is simply connected. The formulation of Theorem B given here, and the proof of the theorem and its converse in the more general case, are due to Walsh,[6] who has also generalized Theorem A and its converse to the case of any finite number of mutually exterior closed Jordan regions.[7]

In the proof of these theorems, Kalmár and Walsh make use of the uniform convergence on C of the sequence $\{L_n(z; f)\}$ only for functions $f(z)$ belonging to a certain subclass A_C' of A_C. This subclass A_C' is the class of all functions of the type $f(z) = 1/(t - z)$, where the parameter t assumes the affix of every point in K. It is our main purpose in this paper to show that the conclusions of Theorem A and Theorem B (with certain restrictions on the set C in the latter case) can be derived from an initial assumption which is much weaker than the condition of uniform convergence of the sequence $\{L_n(z; f)\}$ for each $f(z) \in A_C'$. In fact, our hypothesis will be merely that at a single interior point z_0 of C, the following inequality holds:

$$\overline{\lim_{n\to\infty}} |L_n(z_0; f)|^{1/n} \leqq 1, \qquad f(z) \in A_C'. \tag{1.1}$$

[4] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, Math. Ann., vol. 77 (1916), pp. 313-352.

[5] Kalmár, loc. cit.

[6] Walsh, op. cit., pp. 154-155, 159-162.

[7] Walsh, op. cit., pp. 168-170. See also G. Szegö, Bemerkungen zu einer Arbeit von Herrn M. Fekete, Math. Zeit., vol. 21 (1924), pp. 203-208.

In the course of our discussion, we shall derive several auxiliary results which appear to be of some interest in themselves, and which will accordingly be developed in some detail They include several new non-trivial sufficient conditions for the convergence of the sequence $\{L_n(z; f)\}$ (Theorems 2.2, 4.1, 5.1, 5.2), an inequality which promises to be useful in the study of the fundamental polynomials $\omega_n(z) = \prod_{k=1}^{n} (z - \alpha_k^{(n)})$ (§3), and a new sufficient condition that points be uniformly-densely distributed on a Jordan curve (Theorem 5.1). In connection with the inequality, we shall digress briefly to discuss a recent paper by Lammel[8] (§4).

2. Preliminary Theorems

We shall show in this section that condition (1.1), without the restriction that z_0 be an interior point of C, implies the convergence of the sequence $\{L_n(z_0; f)\}$ for any $f(z) \in A_C$.

THEOREM 2.1. *Let C be an arbitrary closed limited point set whose complement K is connected. Let the limit points of M all lie on C. Let z_0 be any point of C and let K' be an arbitrary closed limited point set interior to K. If*

$$\overline{\lim_{n\to\infty}} \, |L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K, \tag{2.1}$$

then

$$\overline{\lim_{n\to\infty}} \left| \frac{\omega_n(z_0)}{\omega_n(t)} \right|^{1/n} \leqq \rho < 1, \qquad t \in K', \tag{2.2}$$

where ρ is independent of t.

The proof is indirect. For a fixed t, if n is sufficiently large, we may write

$$L_n[z; 1/(t - z)] = \frac{1}{t - z}\left[1 - \frac{\omega_n(z)}{\omega_n(t)}\right], \qquad z \neq t. \tag{2.3}$$

Suppose now that for some $t_0 \in K$,

$$\overline{\lim_{n\to\infty}} \left| \frac{\omega_n(z_0)}{\omega_n(t_0)} \right|^{1/n} = \mu_0 \geqq 1. \tag{2.4}$$

If $\mu_0 > 1$, we obtain a contradiction at once by reference to (2.1) and (2.3). But in any case we may proceed as follows:

The functions $(\log |\omega_n(t)|)/n$ form a normal family of harmonic functions in any closed limited sub-region of K, provided that n is sufficiently large, because in that case $|t - \alpha_k^{(n)}|$, $k = 1, \cdots, n$, is not only uniformly bounded, but also uniformly bounded from zero. Since $(\log |\omega_n(t)|)/n \to \infty$ as $t \to \infty$, no limit function of the family can be identically constant in K. Let the subscript n' run through an infinite subsequence of the positive integers such that if n in (2.4)

[8] E. Lammel, Über Approximation regulärer Funktionen eines komplexen Argumentes durch rationale Funktionen, Math. Zeit., vol. 46 (1940), pp. 104–116.

is replaced by n', the symbol $\overline{\lim}$ can be replaced by the symbol lim. There exists a limit function $W(t)$ defined by the equation

$$\log W(t) = \lim_{n''\to\infty} \frac{\log |\omega_{n''}(t)|}{n''}, \qquad t \in K, t \neq \infty,$$

where n'' assumes an infinite subsequence of the values assumed by n'. The Principle of the Maximum for harmonic functions, applied to $\log W(t)$, now shows that there exists a point $t_1 \in K$ such that $W(t_1) < W(t_0)$. We have

$$\lim_{n''\to\infty} \left|\frac{\omega_{n''}(z_0)}{\omega_{n''}(t_1)}\right|^{1/n''} = \lim_{n''\to\infty} \left|\frac{\omega_{n''}(z_0)\omega_{n''}(t_0)}{\omega_{n''}(t_0)\omega_{n''}(t_1)}\right|^{1/n''} = \mu_0 \frac{W(t_0)}{W(t_1)} > 1.$$

Thus there exists a number $\mu_1 > 1$ such that for $n'' > n(\mu_1)$, we can write the inequality

$$\left|\frac{\omega_{n''}(z_0)}{\omega_{n''}(t_1)}\right| \geqq \mu_1^{n''}.$$

Substituting into (2.3), we find that

$$|L_{n''}[z_0; 1/(t_1 - z_0)]|^{1/n''} \geqq \left|\frac{1}{t_1 - z_0}\right|^{1/n''} (\mu_1^{n''} - 1)^{1/n''}, \qquad n'' > n(\mu_1),$$

which contradicts (2.1). Thus for each t in K the first member of (2.2) is less than unity. The fact that it is bounded from unity for $t \in K'$ now follows from the observation that if n is sufficiently large, the functions whose superior limit appears in (2.2) form a bounded equicontinuous family for $t \in K'$, so the superior limit is a continuous function of t on K'.

THEOREM 2.2. *Let C be an arbitrary closed limited point set whose complement K is connected. Let the limit points of M all lie on C. A necessary and sufficient condition that*

$$(2.5) \qquad L_n(z_0; f) - f(z_0) = 0(r^n), \qquad r < 1, \qquad z_0 \in C, \qquad f(z) \in A_C,$$

where r depends on $f(z)$, is that

$$\overline{\lim_{n\to\infty}} |L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K.$$

The necessity is obvious from (2.3). In the sufficiency, we use the fact that given a function $f(z) \in A_C$, there exists a point set S consisting of a finite number of regions, each bounded by a finite number of non-intersecting rectifiable curves, such that S contains C in its interior and such that $f(z)$ is analytic and single-valued on the corresponding closed set $\bar{S}$.[9] Let the set K' of Theorem 2.1 be the boundary of S. The Cauchy-Hermite formula for $L_n(z; f)$ can be set up as follows:

$$L_n(z_0; f) = \frac{1}{2\pi i} \int_{K'} \frac{f(t)}{t - z} \left[1 - \frac{\omega_n(z_0)}{\omega_n(t)}\right] dt, \qquad n > n(f),$$

and the remainder of the proof follows at once from Theorem 2.1.

[9] Walsh, op. cit., pp. 12–13.

We add three remarks concerning these theorems.

(a) It is not difficult to devise sequences M for which (2.1) obviously holds. If the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$, are assumed to be distinct, we have the familiar formula

$$L_n(z_0; f) = \sum_{k=1}^{n} f(\alpha_k^{(n)}) \frac{\omega_n(z_0)}{(z_0 - \alpha_k^{(n)})\omega_n'(\alpha_k^{(n)})}.$$

Thus a sufficient condition that the sequence M should satisfy (2.1) is that

$$\left| \frac{\omega_n(z_0)}{(z_0 - \alpha_k^{(n)})\omega_n'(\alpha_k^{(n)})} \right| \leqq \chi_n, \qquad k = 1, \cdots n, \qquad \overline{\lim_{n\to\infty}} \chi_n^{1/n} \leqq 1. \tag{2.6}$$

Sequences M which satisfy (2.6) for every point $z_0 \in C$ are easily shown to exist, and have been studied by a number of authors; an example of such a sequence M is given by the "Fekete points."[10]

(b) If instead of being valid merely at a single point z_0, (2.1) or (2.2) holds at each point z of a closed subset C' of C containing no limit points of M, then the functions whose superior limit appears in (2.2) form (for n sufficiently large) a bounded equicontinuous family in the two variables z and t, $z \in C'$, $t \in K'$. Accordingly, in this case the number ρ in (2.2) may be chosen so as to be independent of both z and t, $z \in C'$, $t \in K'$, and (2.5) holds uniformly for $z \in C'$.

(c) The proof of Theorem 2.1 does not apply to the case in which K is not connected, because then it may happen that the limit function $W(t)$ used in the proof is a constant. In fact, both Theorems 2.1 and 2.2 are false in this case, as may be seen by letting C be a circular annulus whose outer boundary is the unit circle, and letting the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$ be the n-th roots of unity.

3. An inequality

LEMMA 3.1. *Let $a_1, a_2, \cdots, a_n$ be n real numbers such that $q\eta \leqq a_k \leqq 1$, $k = 1, \cdots, n$, where $0 < \eta < 1$, $0 < q < 1/\eta$. Then*

$$q(a_1, a_2, \cdots, a_n) = \prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n}$$

$$\leqq \begin{cases} \eta, & \eta^2 + \eta \geqq 1/(q+1) \\ (q\eta)^{\kappa} \dfrac{\log q - \log (q+1) - \log (\eta+1)}{\log \eta + \log (q+1) + \log (\eta+1)}, & \eta^2 + \eta < 1/(q+1), \end{cases} \tag{3.1}$$

where

$$\kappa = \frac{\log \log q\eta - \log [\log \eta + \log (q+1) + \log (\eta+1)] - \log (\eta+1)}{\log (q+1) - \log q - \log (\eta+1)}.$$

It should be observed that the right member of the inequality is independent of n.

[10] See Walsh, op. cit., pp. 170–173.

To prove the inequality, we first remark that the maximum value of the function $g(a_1, a_2, \cdots, a_n)$ in the n-dimensional cube $q\eta \leqq a_k \leqq 1, k = 1, \cdots, n$, is assumed at one of the "corners" of this cube. This may be readily established by using the symmetry of g and by examining the partial derivatives of the function $g(x_1, x_2, \cdots, x_k, a_{k+1}, \cdots, a_n)$ with respect to the a's, where each of the x's assumes either the value $q\eta$ or 1. We leave the details to the reader.

Suppose now that each a is an x, and that m of the x's are equal to $q\eta$ and $n - m$ of them are equal to 1. Then

$$q(x_1, x_2, \cdots, x_n) = h(m) = (1 + \eta)^{(n-m)/n}(q\eta + \eta)^{m/n} - (q\eta)^{m/n}.$$

Let the range of variation of m be extended to include all the real numbers. Computing the derivative of $h(m)$ by the methods of elementary calculus, we find at once that if $\log (1 + \eta) + \log (q\eta + \eta) \geqq 0$, then $h'(m) > 0$ for all m, so in this case the maximum of $h(m)$ in the interval $0 \leqq m \leqq n$ occurs when $n = m$. Since $h(n) = \eta$, we obtain the first part of (3.1). On the other hand, if $\log (1 + \eta) + \log (q\eta + \eta) < 0$, then there exists a unique root $m = \hat{m}$ of the equation $h'(m) = 0$ in the interval $0 < m < n$, and $\hat{m}$ is readily found to be the maximum point of $h(m)$. The rather artificial-looking expression in the second part of (3.1) is simply $h(\hat{m})$, and $\kappa = \hat{m}/n$.

It is apparent from the method of proof and from the order of magnitude of $h(m) - h([m])$ (where $[m]$ denotes the largest integer in m) that the inequality cannot be essentially improved. The inequality is more convenient when stated in less explicit form. Since κ is positive, the second expression in the third member of (3.1) is clearly dominated by $[\log 2 + \log (1 + 1/q)]/|\log \eta|$. Thus we obtain the following corollary:

LEMMA 3.2. *Let $a_1, \cdots, a_n$, η, and q satisfy the conditions in Lemma* 3.1. *Then there exists a number A_q independent of n and η such that*

$$\prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n} \leqq \frac{A_q}{|\log \eta|}. \tag{3.2}$$

It happens that the right member of (3.2) reflects accurately the order of magnitude of the right member of (3.1) for small values of η, because $\lim_{\eta \to 0} (q\eta)^{\kappa} = 1/e$.

We now apply these results to the fundamental polynomials $\omega_n(z)$.

THEOREM 3.1. *Let z be a point of the complex z-plane such that $z \neq \alpha_k^{(n)}$, $k = 1, \cdots, n$. Let D be a number such that $|z - \alpha_k^{(n)}| \leqq D$, $k = 1, \cdots, n$. If t is a point which satisfies the inequality $|t - z| \leqq \delta$, where $0 < \delta < D$, and if q be chosen so that $q\delta \leqq |z - \alpha_k^{(n)}|$, $k = 1, \cdots, n$, then a number $A_{q,D}$ exists which is independent of n, z, t, and δ, such that*

$$|\omega_n(t)|^{1/n} \leqq |\omega_n(z)|^{1/n} + \frac{A_{q,D}}{|\log \delta|}.$$

We add three remarks concerning these theorems.

(a) It is not difficult to devise sequences M for which (2.1) obviously holds. If the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$, are assumed to be distinct, we have the familiar formula

$$L_n(z_0; f) = \sum_{k=1}^{n} f(\alpha_k^{(n)}) \frac{\omega_n(z_0)}{(z_0 - \alpha_k^{(n)})\omega_n'(\alpha_k^{(n)})}.$$

Thus a sufficient condition that the sequence M should satisfy (2.1) is that

$$\left| \frac{\omega_n(z_0)}{(z_0 - \alpha_k^{(n)})\omega_n'(\alpha_k^{(n)})} \right| \leqq \chi_n, \qquad k = 1, \cdots n, \qquad \overline{\lim_{n\to\infty}} \chi_n^{1/n} \leqq 1. \tag{2.6}$$

Sequences M which satisfy (2.6) for every point $z_0 \in C$ are easily shown to exist, and have been studied by a number of authors; an example of such a sequence M is given by the "Fekete points."[10]

(b) If instead of being valid merely at a single point z_0, (2.1) or (2.2) holds at each point z of a closed subset C' of C containing no limit points of M, then the functions whose superior limit appears in (2.2) form (for n sufficiently large) a bounded equicontinuous family in the two variables z and t, $z \in C'$, $t \in K'$. Accordingly, in this case the number ρ in (2.2) may be chosen so as to be independent of both z and t, $z \in C'$, $t \in K'$, and (2.5) holds uniformly for $z \in C'$.

(c) The proof of Theorem 2.1 does not apply to the case in which K is not connected, because then it may happen that the limit function $W(t)$ used in the proof is a constant. In fact, both Theorems 2.1 and 2.2 are false in this case, as may be seen by letting C be a circular annulus whose outer boundary is the unit circle, and letting the points $\alpha_k^{(n)}$, $k = 1, \cdots, n$ be the n-th roots of unity.

3. An inequality

LEMMA 3.1. *Let $a_1, a_2, \cdots, a_n$ be n real numbers such that $q\eta \leqq a_k \leqq 1$, $k = 1, \cdots, n$, where $0 < \eta < 1$, $0 < q < 1/\eta$. Then*

$$q(a_1, a_2, \cdots, a_n) = \prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n}$$

$$\leqq \begin{cases} \eta, & \eta^2 + \eta \geqq 1/(q+1) \\ (q\eta)^{\kappa} \dfrac{\log q - \log (q+1) - \log (\eta+1)}{\log \eta + \log (q+1) + \log (\eta+1)}, & \eta^2 + \eta < 1/(q+1), \end{cases} \tag{3.1}$$

where

$$\kappa = \frac{\log \log q\eta - \log [\log \eta + \log (q+1) + \log (\eta+1)] - \log (\eta+1)}{\log (q+1) - \log q - \log (\eta+1)}.$$

It should be observed that the right member of the inequality is independent of n.

[10] See Walsh, op. cit., pp. 170–173.

To prove the inequality, we first remark that the maximum value of the function $g(a_1, a_2, \cdots, a_n)$ in the n-dimensional cube $q\eta \leqq a_k \leqq 1, k = 1, \cdots, n$, is assumed at one of the "corners" of this cube. This may be readily established by using the symmetry of g and by examining the partial derivatives of the function $g(x_1, x_2, \cdots, x_k, a_{k+1}, \cdots, a_n)$ with respect to the a's, where each of the x's assumes either the value $q\eta$ or 1. We leave the details to the reader.

Suppose now that each a is an x, and that m of the x's are equal to $q\eta$ and $n - m$ of them are equal to 1. Then

$$q(x_1, x_2, \cdots, x_n) = h(m) = (1 + \eta)^{(n-m)/n}(q\eta + \eta)^{m/n} - (q\eta)^{m/n}.$$

Let the range of variation of m be extended to include all the real numbers. Computing the derivative of $h(m)$ by the methods of elementary calculus, we find at once that if $\log (1 + \eta) + \log (q\eta + \eta) \geqq 0$, then $h'(m) > 0$ for all m, so in this case the maximum of $h(m)$ in the interval $0 \leqq m \leqq n$ occurs when $n = m$. Since $h(n) = \eta$, we obtain the first part of (3.1). On the other hand, if $\log (1 + \eta) + \log (q\eta + \eta) < 0$, then there exists a unique root $m = \hat{m}$ of the equation $h'(m) = 0$ in the interval $0 < m < n$, and $\hat{m}$ is readily found to be the maximum point of $h(m)$. The rather artificial-looking expression in the second part of (3.1) is simply $h(\hat{m})$, and $\kappa = \hat{m}/n$.

It is apparent from the method of proof and from the order of magnitude of $h(m) - h([m])$ (where $[m]$ denotes the largest integer in m) that the inequality cannot be essentially improved. The inequality is more convenient when stated in less explicit form. Since κ is positive, the second expression in the third member of (3.1) is clearly dominated by $[\log 2 + \log (1 + 1/q)]/|\log \eta|$. Thus we obtain the following corollary:

LEMMA 3.2. *Let $a_1, \cdots, a_n$, η, and q satisfy the conditions in Lemma 3.1. Then there exists a number A_q independent of n and η such that*

$$\prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n} \leqq \frac{A_q}{|\log \eta|}. \tag{3.2}$$

It happens that the right member of (3.2) reflects accurately the order of magnitude of the right member of (3.1) for small values of η, because $\lim_{\eta \to 0} (q\eta)^{\kappa} = 1/e$.

We now apply these results to the fundamental polynomials $\omega_n(z)$.

THEOREM 3.1. *Let z be a point of the complex z-plane such that $z \neq \alpha_k^{(n)}$, $k = 1, \cdots, n$. Let D be a number such that $|z - \alpha_k^{(n)}| \leqq D$, $k = 1, \cdots, n$. If t is a point which satisfies the inequality $|t - z| \leqq \delta$, where $0 < \delta < D$, and if q be chosen so that $q\delta \leqq |z - \alpha_k^{(n)}|$, $k = 1, \cdots, n$, then a number $A_{q,D}$ exists which is independent of n, z, t, and δ, such that*

$$|\omega_n(t)|^{1/n} \leqq |\omega_n(z)|^{1/n} + \frac{A_{q,D}}{|\log \delta|}.$$

For the proof, let $a_k = |z - \alpha_k^{(n)}|/D$, $k = 1, \cdots, n$, and $\eta = \delta/D$. The numbers a_k, η, and q satisfy the conditions of Lemma 3.1. Since $|t - \alpha| \leqq |z - \alpha| + \delta$, we may write

$$|\omega_n(t)|^{1/n} = \prod_{k=1}^{n} |t - \alpha_k^{(n)}|^{1/n} \leqq \prod_{k=1}^{n} (|z - \alpha_k^{(n)}| + \delta)^{1/n}$$

$$= |\omega_n(z)|^{1/n} + D\left[\prod_{k=1}^{n} (a_k + \eta)^{1/n} - \prod_{k=1}^{n} a_k^{1/n}\right].$$

The conclusion of the theorem now follows at once from Lemma 3.2.

THEOREM 3.2. *Let C be an arbitrary closed point set of the extended plane (not the entire plane) with interior points, and let the derivative of the set M be a limited set contained in the boundary of C. Let $\{z_j\}$ be an infinite sequence of finite interior points of C approaching a finite boundary point β of C. Then given any $\epsilon > 0$, there exists a corresponding sequence $\{t_j\}$ of points of the complement of C approaching β, such that for all $j > j(\epsilon)$,*

$$|\omega_n(t_j)|^{1/n} \leqq |\omega_n(z_j)|^{1/n} + \epsilon, \qquad n > n(j). \tag{3.3}$$

For let $D = \overline{\text{bound}}\,|z_j - \alpha_k^{(n)}|$, $j = 1, 2, \cdots$, $k = 1, 2, \cdots, n$, $n = 1, 2, \cdots$. Let $3\epsilon_j/4$ denote the distance from z_j to the boundary of C. By choosing j sufficiently large, say $j > j(\epsilon)$, we can make ϵ_j so small that (in the notation of Theorem 3.1) $A_{1/2,D}/|\log \epsilon_j| < \epsilon$. There exists for each j a point t_j in the complement of C such that $|t_j - z_j| < \epsilon_j$. Moreover, for n sufficiently large, say $n > n(j)$, $|z_j - \alpha_k^{(n)}| \geqq \frac{1}{2}\epsilon_j$, $k = 1, \cdots, n$. The conditions of Theorem 3.1 with $q = \frac{1}{2}$, $\delta = \epsilon_j$, are satisfied for each $j > j(\epsilon)$, provided that $n > n(j)$, and accordingly we obtain (3.3).

In general, t_j and z_j cannot be interchanged in (3.3); that is, given a sequence $\{z_j\}$ in C approaching a boundary point of C, we cannot conclude that there exists a sequence $\{t_j\}$ in the complement of C such that for $j > j(\epsilon)$,

$$||\omega_n(z_j)|^{1/n} - |\omega_n(t_j)|^{1/n}| \leqq \epsilon, \qquad n > n(j), \tag{3.4}$$

where $\epsilon > 0$ is arbitrary. That this situation obtains can easily be proved by examples; it is attributable to the fact that an inequality such as (3.1), where the right member approaches zero with η, is not available if $q = 0$. However, if C is a closed Jordan region bounded by a suitably restricted curve (the requirements fall considerably short of analyticity), then the equicontinuity property of the sequence $\{|\omega_n(z)|^{1/n}\}$ expressed by (3.4) can be established by means of Theorem 3.1. We shall not need this equicontinuity property in the investigation of this sequence which follows, so we shall not discuss the matter further.

4. Applications of the inequality

Let C be a closed limited point set whose complement K is connected; furthermore, let us assume that C has interior points and that all limit points of the

set M lie on the boundary B of C. Suppose now that (1.1) holds at an interior point z_0 of C. There exists a maximal region R_0 of C containing z_0; the boundary of R_0 is contained in B. If n is sufficiently large, the functions $(\log |\omega_n(z)|)/n$ form a normal family of harmonic functions in any closed subregion of R_0. It is a simple consequence of Theorem 2.1 and 3.2 that each limit function of this family, or of the family $|\omega_n(z)|^{1/n}$, is a constant, as we now shall show.

Let $W(z)$ be an arbitrary limit function of the family $|\omega_n(z)|^{1/n}$; $W(z)$ is defined by an equation of the following type:

$$\log W(z) = \lim_{n' \to \infty} \frac{\log |\omega_{n'}(z)|}{n'}, \qquad z \in R_0,$$

where n' runs through some infinite subsequence of the positive integers. If $W(z)$ is not identically constant in R_0, then by the Principle of the Maximum for harmonic functions, there is an infinite sequence $\{z_j\}$ of points of R_0 approaching a point of B such that for some $\epsilon > 0$, $W(z_j) < W(z_0) - 2\epsilon$, $j = 1, 2, \cdots$. But by Theorem 3.2, there exists a subscript J and a point t_J in K such that

$$|\omega_n(t_J)|^{1/n} \leqq |\omega_n(z_J)|^{1/n} + \epsilon, \qquad n > n(J).$$

Therefore

$$\overline{\lim_{n' \to \infty}} |\omega_{n'}(t_J)|^{1/n'} \leqq \lim_{n' \to \infty} |\omega_{n'}(z_J)|^{1/n'} + \epsilon = W(z_J) + \epsilon < W(z_0) - \epsilon;$$

$$\overline{\lim_{n' \to \infty}} |\omega_{n'}(t_J)|^{1/n'} < \lim_{n' \to \infty} |\omega_{n'}(z_0)|^{1/n'}.$$

But the last inequality contradicts (2.2), so $W(z)$ must be a constant. We are now in a position to state the following theorem:

THEOREM 4.1. *Let C be a closed limited point set whose complement K is connected. Let all limit points of the set M lie on the boundary of C. Let z_0 be some interior point of C, and let R_0 be the maximal region of C containing z_0. If*

$$\overline{\lim_{n \to \infty}} |L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K,$$

then

(a) *each limit function in R_0 of the normal family $(\log |\omega_n(z)|)/n$ is a constant;*

(b) *for every function $f(z) \in A_C$, $L_n(z; f) - f(z) = 0(r^n)$, $r < 1$, uniformly on any closed point set interior to R_0.*

Part (b) of the theorem is a simple consequence of part (a) of the theorem, and of Theorems 2.1, 2.2, and Remark (b) of §1. We leave the proof to the reader.

The theorem is false if limit points of M are allowed to lie inside R_0. For

example, if C is the set $|z| \leqq 1$, and if $\alpha_k^{(2n)} = e^{2\pi ik/n}$, $k = 1, \cdots, n$, $\alpha_k^{(2n)} = \frac{1}{2}$, $k = n+1, \cdots, 2n$, then

$$\lim_{n\to\infty} |\omega_{2n}(z)|^{1/2n} = \begin{cases} |z - \frac{1}{2}|^{\frac{1}{2}}, & |z| < 1 \\ (|z|\,|z - \frac{1}{2}|)^{\frac{1}{2}}, & |z| > 1; \end{cases}$$

$$\lim_{n\to\infty} L_{2n}(z; f) = f(z), \quad |z - \tfrac{1}{2}| \leqq \tfrac{1}{2}, f(z) \in A_C;$$

but the sequence $\{L_{2n}(z; f)\}$, $f(z) \in A_C$, may diverge if $|z - \frac{1}{2}| > \frac{1}{2}$.

The problem involved in the proof of part (a) of the theorem is typical of a class of problems arising in the study of interpolation by rational functions in the complex domain, in which certain properties of the sequence $\{|\omega_n(z)|^{1/n}\}$ are to be derived from other properties of this sequence known for z both interior and exterior to a closed set C. In dealing with such problems, the inequality (3.3) may provide a useful tool.[11] The writer intends to discuss this matter elsewhere in more detail. However, it seems not out of place to mention an application of the inequality to a line of investigation which is related to Theorem A, but which is concerned with interpolatory series of rational functions with poles at the points M.

Recently Lammel[12] has derived necessary and sufficient conditions that a certain expansion of the type just described should converge interior to a Jordan region R to any function $f(z) \in A_{\bar{R}}$, where $\bar{R}$ is the corresponding closed region. The points M are chosen on the boundary of R. In studying the necessity, his starting point essentially is the condition

$$\overline{\lim_{n\to\infty}} \frac{|z - a|}{|\omega_n(z)|^{1/n}} \leqq \lim_{n\to\infty} \frac{|t - a|}{|\omega_n(t)|^{1/n}}, \tag{4.1}$$

which is to hold for all $z \in R$ and all t in the complement K of $\bar{R}$, where a is some point of R. From this inequality he infers, by a rather complicated argument, that

$$\lim_{n\to\infty} \frac{|t - a|}{|\omega_n(t)|^{1/n}} = 1, \tag{4.2}$$

uniformly on any closed subset of K. But a line of reasoning similar to that used to establish Theorem 4.1a (the reader will have no trouble in supplying the details) will show that (4.2) is quite an immediate consequence of (4.1) and Theorem 3.2. Moreover in deriving (4.2) by this method, the restriction to a Jordan region becomes quite superfluous; indeed, R may be any region whose boundary is also the boundary of an infinite region.[13]

[11] Another application of the inequality similar to the one in this section appears in the proof of Theorem 5.1b below.

[12] E. Lammel, loc. cit.

[13] It follows from this that Lammel's Theorem 1 (loc. cit., p. 105) is true for such regions R.

5. The revised version of Theorems A and B

We shall now prove Theorem A under the revised hypothesis (1.1), and also Theorem B under this hypothesis, but with the necessary additional restriction on C that it be the closure of a region R whose boundary is also the boundary of an infinite region. Our method of approach is indicated in the following theorem:

THEOREM 5.1. *Let R be a limited region whose boundary B is the boundary of an infinite region K. Let all limit points of the set M lie on B. Necessary conditions that every limit function in R of the normal family* $(\log |\omega_n(z)|)/n$ *should be a constant are that*

(5.1) (a) $\lim_{n\to\infty} |\omega_n(z)|^{1/n} = \Delta$, *uniformly on any closed point set interior to R, where Δ is the transfinite diameter of $R + B$,*

(5.2) (b) $\lim_{n\to\infty} |\omega_n(z)|^{1/n} = \Delta\,|\varphi(z)|$, *uniformly on any closed limited point set interior to K, where $w = \varphi(z)$ maps K conformally onto the region $|w| > 1$ so that the points at infinity correspond to each other.*

If R is a Jordan region and if all the points M lie on B, then a further necessary condition is that the points M are uniformly-densely distributed on B.

Put in another way, the theorem states that if the mean radii of the lemniscates $|\omega_n(z)| = \mu_n$, $n = 1, 2, \cdots$, passing through an arbitrary point $z \in R$, approaches as $n \to \infty$ a limit which is constant for all $z \in R$, then this limit must be the same as the limit of the mean radii of the minimal covering lemniscates of $R + B$ of respective degrees n.[14]

Let $\zeta_n(z) = (\log |\omega_n(z)|)/n$. Let $\{\zeta_{n'}(z)\}$ be an arbitrary convergent subsequence of the family $\zeta_n(z)$. The convergence of this subsequence is uniform in some neighborhood N of a point $z_1 \in R$, and the neighborhood is independent of the particular subsequence under consideration. Since the limit of the subsequence is a constant, the two corresponding subsequences of partial derivatives with respect to x and y, $z = x + iy$, both converge uniformly to zero in N.[15] Therefore

$$\lim_{n'\to\infty} \frac{\sum_1^{n'} (z - \alpha_k^{n'})^{-1}}{n'} = 0, \quad \text{uniformly,} \quad z \in N. \tag{5.3}$$

But the sequence $\{\zeta_{n'}(z)\}$ was arbitrary, so we are able to conclude that every limit function of the normal family of analytic functions $[\sum_1^n (z - \alpha_k^{(n)})^{-1}]/n$ is identically zero in N. Therefore (5.3) is still true if n' is replaced by n.

---

[14] The latter limit is known to be Δ. See M. Fekete, Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen, etc., Math. Zeit., vol. 17 (1923), pp. 228–249.

[15] W. F. Osgood, Lehrbuch der Funktionentheorie, fifth edition, Leipzig, 1928, vol. I, p. 682.

Differentiating and using the Weierstrass Double Series Theorem, we obtain the equations

$$\lim_{n\to\infty} \mathcal{R}\left[\frac{A_\nu \sum_{k=1}^{n} (z_1 - \alpha_k^{(n)})^{-\nu}}{n}\right] = 0, \qquad \nu = 1, 2, 3, \cdots, \tag{5.4}$$

where the numbers A_ν are any complex constants.

Let $z = \psi(w)$ be the inverse of the mapping function $w = \varphi(z)$. The function $\psi(w)$ is analytic and single-valued for $|w| > 1$ and may be represented by a Laurent series of the following type:

$$z = \psi(w) = cw + \sum_{j}^{\infty} C_k w^{-k}, \qquad c \neq 0, \qquad |w| > 1,$$

where $|c| = \Delta$.[16] Thus

$$\lim_{z\to\infty} \left|\frac{z}{\varphi(z)}\right| = \left|\frac{1}{\varphi'(\infty)}\right| = \Delta. \tag{5.5}$$

Therefore the function $l(z) = \log |(z_1 - z)/\varphi(z)|$, when assigned the value $\log \Delta$ at infinity, is harmonic for $z \in K$; and if $|\varphi(z)|$ is defined to be unity on B, $l(z)$ is continuous in $K + B$. Invert the plane in the point z_1 by means of the transformation $z' = 1/(z_1 - z)$, and let K' be the image of K. The function $l_1(z') = l(z_1 - 1/z')$ is harmonic interior to the limited region K' and continuous in the corresponding closed region $\bar{K}'$. A theorem of Walsh[17] now states that given an arbitrary $\epsilon > 0$, there exists a harmonic polynomial $P(z')$ such that $|l_1(z') - P(z')| < \epsilon/4$, $z' \in \bar{K}'$. The polynomial $P(z')$ is an expression of the type $A_0 + \sum_1^m \mathcal{R}(A_\nu z'^{\nu})$, where A_0 is a real constant and the A_ν are complex constants. In particular, we find that $|l_1(0) - P(0)| = |\log \Delta - A_0| < \epsilon/4$. If we use this fact when we reverse the inversion, we obtain the inequality

$$\left|l(z) - \log \Delta - \sum_{\nu=1}^{m} \mathcal{R}[A_\nu (z_1 - z)^{-\nu}]\right| < \frac{\epsilon}{2}, \qquad z \in K + B,$$

or

$$\left|\log |z_1 - z| - \log \Delta - \sum_{\nu=1}^{m} \mathcal{R}[A_\nu (z_1 - z)^{-\nu}]\right| < \frac{\epsilon}{2}, \qquad z \in B.$$

Therefore

$$\left|\frac{1}{n}\sum_{k=1}^{n} \log |z_1 - \alpha_k^{(n)}| - \log \Delta - \frac{1}{n}\sum_{\nu=1}^{m} \mathcal{R}\left[A_\nu \sum_{k=1}^{n} (z_1 - \alpha_k^{(n)})^{-\nu}\right]\right| < \frac{\epsilon}{2},$$

$$n = 1, 2, \cdots.$$

---

[16] Fekete, loc. cit., pp. 237–240.

[17] J. L. Walsh, The approximation of harmonic functions by harmonic polynomials and by harmonic rational functions, Bull. Am. Math. Soc., vol. 35 (1929), pp. 499–544; p. 503.

Referring now to (5.4), we find that for all values of n sufficiently large, $|(\log|\omega_n(z_1)|)/n - \log\Delta| < \epsilon$. Thus all the constant limit functions of the family $(\log|\omega_n(z)|)/n$ have the value $\log\Delta$, which means that $\lim_{n\to\infty}(\log|\omega_n(z)|)/n = \log\Delta$, $z \in R$. Since the functions form a normal family in R, this limit is uniform on any closed point set interior to R. We have established part (a) of the theorem.

A suitably chosen branch of the function $[\omega_n(z)]^{1/n}/(\Delta\cdot\varphi(z))$, when defined to have the value unity at infinity (in accordance with (5.5)), is analytic in any closed subregion of K if n is sufficiently large, and is uniformly bounded in modulus in K, the bound being independent of n. Thus the functions $[\omega_n(z)]^{1/n}/(\Delta\varphi(z))$ form a normal family of analytic functions in any closed subregion of K, if n is sufficiently large. Let $\theta(z)$ be a limit function of this family. Since $|\theta(\infty)| = 1$, it follows from the Principle of the Maximum that either $|\theta(z)| \equiv 1$, or else there exists an infinite sequence $\{t_j\}$ of points of K approaching a point of B such that $\lim_{j\to\infty}|\theta(t_j)| < 1$. But Theorem 3.2 and part (a) of the present theorem show that the latter case is impossible. For if such a sequence $\{t_j\}$ existed, then (since $\lim_{j\to\infty}\Delta|\varphi(t_j)||\theta(t_j)| < \Delta$) there would exist a number $\epsilon > 0$ such that $|\omega_{n'}(t_j)|^{1/n'} < \Delta - 2\epsilon$, $n' \geqq n'(j)$, $j > j(\epsilon)$, where n' runs through some infinite subsequence of the positive integers. However, by Theorem 3.2, there exists a subscript $J > j(\epsilon)$ and a point $z_J \in R$ such that

$$|\omega_{n'}(z_J)|^{1/n'} \leqq |\omega_{n'}(t_J)|^{1/n'} + \epsilon < \Delta - 2\epsilon + \epsilon, \qquad n' > n'(J).$$

If we apply part (a) of the present theorem to the first member of this inequality, we obtain a contradiction at once. Thus every limit function of the normal family $[\omega_n(z)]^{1/n}/(\Delta\varphi(z))$ is a constant of modulus unity. This establishes part b of the theorem.

The proof of the last part of the theorem is now a consequence of the work of Kalmár,[18] who showed in his proof of Theorems A and B that if R is a Jordan region and if the points M lie on B, then equation (5.2) implies that the points M are uniformly-densely distributed on B.

Combining Theorems 4.1 and 5.1, we may state our revised version of Theorems A and B very briefly as follows:

THEOREM 5.2. *Under the hypotheses of Theorem* 5.1 *concerning R and M, the conditions* (a) *and* (b) *of that theorem and the uniformly dense distribution of the points M are necessary conditions that at some point $z_0 \in R$,*

$$\overline{\lim_{n\to\infty}}\,|L_n[z_0; 1/(t - z_0)]|^{1/n} \leqq 1, \qquad t \in K. \tag{5.6}$$

Of course these necessary conditions are also sufficient that (5.6) hold. In-

[18] Kalmár, loc. cit. See also Walsh, op. cit., pp. 168–170.

deed, in the light of Theorem B, it is seen that (5.6) or the hypotheses of Theorem 5.1 are sufficient conditions that

$$\lim_{n\to\infty} L_n(z;f) = f(z), \qquad z \in R + B, \qquad f(z) \in A_{R+B},$$

the convergence being not only uniform with respect to z, but "maximal."[19]

The conclusions of Theorem 5.2 would be false if z_0 were a boundary point of R, or if $R + B$ were replaced by any more general closed set of C of which z_0 is an interior point, even if the complement of C were simply connected. The reader will have no trouble verifying these statements by means of trivial examples similar to the one in §4.

It is perhaps of interest to observe that an alternative method of proof for Theorem 5.1 is available which in no way depends on Kalmár's work. The procedure consists in using a well-known necessary and sufficient condition of Weyl[20] that the points $\varphi(\alpha_k^{(n)})$ be uniformly-densely distributed on the unit circle, together with equations (5.4), to establish the last part of the theorem first. Equations (5.1) and (5.2) for a Jordan region may then be obtained readily from the Gauss Mean Value Theorem. The validity of these equations can then be extended to the more general case by approximating the region R by means of its Green's level curves.

CORNELL UNIVERSITY.

[19] Walsh, op. cit., p. 154.

[20] Weyl, loc. cit., pp. 313–316.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

HILBERT DISTANCES AND POSITIVE DEFINITE FUNCTIONS

BY S. BOCHNER

(Received January 20, 1941)

I. HILBERT DISTANCES

If $\mathfrak{S}$ is any topological space then a *continuous* complex function of two variables $f(P, Q)$ on $\mathfrak{S} \times \mathfrak{S}$ shall belong to class $\mathfrak{P}$ (positive definite functions) if

1) $f(P, Q) = \overline{f(Q, P)}$

2) $f(P, P) = f(Q, Q)$

and

3′) for any integer n, any points $P_1, \cdots, P_n$ and any complex numbers $\rho_1, \cdots, \rho_n$ the relation

$$\sum_{i,j=1}^{n} f(P_i, P_j)\rho_i\overline{\rho_j} \geqq 0, \tag{1}$$

holds.

The wider class $\mathfrak{P}_0$ shall consist of all those functions for which, other assumptions being unchanged, the decisive inequality (1) is assumed to hold only subject to the restriction

$$\sum_{i=1}^{n} \rho_i = 0. \tag{2}$$

Just how much wider the class $\mathfrak{P}_0$ actually is will be established under further assumptions. The *real* classes $\mathfrak{P}$ and $\mathfrak{P}_0$ shall consist of all real-valued functions of these classes.

We will call a function $\rho(P, Q)$ a distance function if $\rho(P, Q) = \rho(Q, P) \geqq 0$, $\rho(P, P) = 0$ and $\rho(P, Q) + \rho(Q, R) \geqq \rho(P, R)$. We call $\rho(P, Q)$ a *proper* distance if $P \neq Q$ implies $\rho(P, Q) > 0$. Introducing the real Hilbert space $\mathfrak{H}$ of sequences $\{x_n\}$, $\{y_n\}$ with the customary distance $\left(\sum_n (x_n - y_n)^2\right)^{1/2}$, we now call $\rho(P, Q)$ a *Hilbert distance* on $\mathfrak{S}$ if it is possible to map $\mathfrak{S}$ into $\mathfrak{H}$ in such a way that the value of $\rho(P, Q)$ shall be equal to the value of the latter distance for the transforms of P and Q. The following decisive criterion has been established by K. Menger and I. J. Schoenberg:[1] *if $\mathfrak{S}$ is separable, then $\rho(P, Q)$ is a Hilbert distance if and only if* $-\rho(P, Q)^2$ belongs to $\mathfrak{P}_0$. Also, Schoenberg discovered the fact that a function $f(P, Q)$ belongs to $\mathfrak{P}_0$ if and only if $e^{\lambda f(P,Q)}$ belongs to $\mathfrak{P}$ for each

[1] See I. J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. 44 (1938), 522–536.

[2] Positive definite functions on non-compact but commutative groups have been analyzed recently in A. Powzner, Doklady U.S.S.R., 28 (1940) 294–295, and D. Raikow, Doklady U.S.S.R., 27 (1940), 324–327 and 28 (1940), 296–300.

positive λ, and he analyzed Hilbert distances on open euclidean and Hilbert spaces.

In the present note we will investigate Hilbert distances for some types of *general* but *compact* separable spaces. We will obtain new relations between the classes $\mathfrak{P}$ and $\mathfrak{P}_0$ and structural criteria in terms of expansions in orthogonal systems.[2]

II. Compact Spaces with Bounded Measure

Since $f(P, Q)$ is continuous, if $\mathfrak{S}$ is a finite interval, say, then condition 3') for $\mathfrak{P}$ can be replaced by the following condition:

3) for each integrable $\rho(P)$ the relation

$$\iint f(P, Q)\rho(P)\rho(Q)\, dP\, dQ \geqq 0 \tag{3}$$

holds;[3] and for $\mathfrak{P}_0$ relation (3) holds subject to the restriction

$$\int \rho(P)\, dP = 0. \tag{4}$$

This replacement of sums by integrals is admissible for more general spaces.

Lemma 1. *Conditions* 3) *and* 3') *are equivalent if* α), $\mathfrak{S}$ *is both separable and compact,* β) *there exists a Lebesque measure* dP *on* $\mathfrak{S}$ *for which* γ) *all Borel sets are measurable,* δ) *every open set has non-vanishing measure and* ϵ) *the total space has measure* 1.

Proof: We first observe that $\mathfrak{S}$ and $\mathfrak{S} \times \mathfrak{S}$ are each bi-compact, and that therefore every continuous function $f(P, Q)$ is uniformly continuous and bounded. In the proof M will be independent of ϵ. We will give the proof of the lemma for functions of $\mathfrak{P}_0$, for functions of $\mathfrak{P}$ it is even simpler. Let $f(P, Q)$ satisfy 3') and let $\rho(P)$ be any function satisfying (4). Given $\epsilon > 0$ we partition $\mathfrak{S}$ into a finite number of Borel sets $\mathfrak{S}_1, \cdots, \mathfrak{S}_n$ such that $f(P, Q)$ oscillates by less than ϵ on each $\mathfrak{S}_i \times \mathfrak{S}_j$. Choosing an arbitrary point P_i in $\mathfrak{S}_i$ and putting $\rho_i = \int_{\mathfrak{S}_i} \rho(P)\, dP$ we obviously have (2) and

$$\left| \sum_{i,j=1}^{n} f(P_i, P_j)\rho_i\overline{\rho_j} - \iint f(P, Q)\rho(P)\overline{\rho(Q)}\, dP\, dQ \right| < \epsilon M. \tag{5}$$

Letting $\epsilon \to 0$ we obtain (3). Conversely, if $f(P, Q)$ satisfies 3), and distinct points $\{P_i\}$ and numbers $\{\rho_i\}$ are given, and if (2) holds, then for $\epsilon > 0$, we pick a set of disjoint neighborhoods $\mathfrak{S}_i$ of P_i, such that $f(P, Q)$ oscillates by less than ϵ on each $\mathfrak{S}_i \times \mathfrak{S}_j$. Putting $\rho(P) = \dfrac{\rho_i}{\text{measure } \mathfrak{S}_i}$ for $P \in \mathfrak{S}_i$, and $\rho(P) = 0$ for $P \in \mathfrak{S} - (\mathfrak{S}_1 + \cdots + \mathfrak{S}_n)$, we have relation (2) and (5), and, by a limit, relation (1).

[3] Whenever the range of integration is not indicated it is the total set to which the variable refers.

In what follows we will always assume that $\mathfrak{S}$ satisfies the assumptions of the lemma.

THEOREM 1. *If $f(P, Q)$ belongs to $\mathfrak{P}$, if $\varphi(P)$ is continuous and c is a real constant, then*

$$g(P, Q) = f(P, Q) + \varphi(P) + \overline{\varphi(Q)} - c \tag{6}$$

belongs to $\mathfrak{P}_0$ except for property 2).

Conversely, if $g(P, Q)$ belongs to $\mathfrak{P}_0$, then there exists a continuous function $\varphi(P)$ and a real constant c such that

$$f(P, Q) = g(P, Q) - \varphi(P) - \overline{\varphi(Q)} + c \tag{7}$$

belongs to $\mathfrak{P}$, except for property 2), as for instance

$$\varphi(P) = \int g(P, Q)\, dQ \tag{8}$$

$$c = \int \varphi(P)\, dP = \iint g(P, Q)\, dP\, dQ. \tag{9}$$

PROOF: The first part follows from the fact that (4) implies

$$\iint \varphi(P)\rho(P)\overline{\rho(Q)}\, dP\, dQ = \int \varphi(P)\rho(P)\, dP \cdot \int \overline{\rho(Q)}\, dQ = 0; \tag{10}$$

the second part follows from the fact that (7), (8), (9) and $\lambda = \int \rho(P)\, dP$ implies

$$\iint f(P, Q)\rho(P)\overline{\rho(Q)}\, dP\, dQ = \iint g(P, Q)(\rho(P) - \lambda)\overline{(\rho(Q) - \lambda)}\, dP\, dQ.$$

We call a function $f(P, Q)$ measure invariant if $\int f(P, Q)\, dQ$ is independent of P. This leads to

THEOREM 2. *A function $g(P, Q)$ belongs to the measure invariant class $\mathfrak{P}_0$, if and only if it can be written in the form*

$$f(P, Q) - f(\Phi, \Phi) \tag{11}$$

where $f(P, Q)$ belongs to the measure invariant class $\mathfrak{P}$ and Φ is any point of $\mathfrak{S}$.

We now assume the existence of a fixed transitive group $\mathfrak{C}$ of continuous transformations of $\mathfrak{S}$ into itself, and we assume that our measure is invariant under the operations of $\mathfrak{C}$. We call a function $f(P, Q)$ group invariant if for all elements s of $\mathfrak{C}$, $f(sP, sQ) = f(P, Q)$, where sP is the transform of the point P under the transformation s. In particular $f(sP, sP) = f(P, P)$. Now if P and Q are any two points, since our group is transitive there exists an element s such that $Q = sP$. Thus from now onward property 2) of $\mathfrak{P}$ will be automatically fulfilled for any function $f(P, Q)$ which is group invariant. In order to

emphasize the independence of $f(P, P)$ of the special point P we will designate it by $f(\Phi, \Phi)$ where Φ is any fixed point which is chosen appropriately. For instance if $\mathfrak{S}$ itself is a group it is customary to take for Φ the identity element of the group.

If $f(P, Q)$ is group invariant we also have

$$\varphi(P) = \int f(P, Q)\, dQ = \int f(sP, sQ)\, dQ = \int f(sP, sQ)\, d(sQ)$$
$$= \int f(sP, Q)\, dQ = \varphi(sP),$$

and this implies that $\varphi(P)$ is a constant. In other words, if $f(P, Q)$ is group invariant it is also measure invariant. This leads to

THEOREM 3. *A function $g(P, Q)$ belongs to the group invariant class $\mathfrak{P}_0$ if and only if it can be written in the form* (11) *where $f(P, Q)$ is group invariant and belongs to $\mathfrak{P}$.*

A function $\rho(P, Q)$ is a group invariant Hilbert distance if and only if it can be represented in the form

$$\sqrt{f(\Phi, \Phi) - f(P, Q)},$$

where $f(P, Q)$ is group invariant and belongs to the real class $\mathfrak{P}$. The function $\rho(P, Q)$ is a proper distance if and only if $f(P, Q) < f(\Phi, \Phi)$ for $P \neq Q$.

III. GROUP INVARIANT FUNCTIONS

In the present section we will supplement Theorem 3 by statements involving expansions in orthonormal systems. An orthonormal system $\{\varphi_n(P)\}$ is of course defined by the property

$$\int \varphi_m(P)\overline{\varphi_n(P)}\, dP = \delta_{mn};$$

the orthonormal system we will consider will automatically consist of *continuous* functions and it will be *complete* both in the space of continuous functions and in the space of integrable functions.

The simplest space $\mathfrak{S}$ is the torus $0 \leqq x < 2\pi$ with the group of translations on it. Any continuous function $f(x, y)$ on $\mathfrak{S} \times \mathfrak{S}$ has a Fourier expansion

$$\sum_{m,n} a_{m,n} e^{imx} e^{-iny}. \tag{12}$$

Group invariance means that $f(x, y) = f(x + s, y + s)$ and this implies

$$a_{m,n} = a_{m,n} e^{i(m-n)s}.$$

Thus, $a_{m,n} = 0$ if $m \neq n$, and (12) has the form

$$\sum_{-\infty}^{\infty} a_m e^{im(x-y)}. \tag{13}$$

Now $f(x, y)$ belongs to $\mathfrak{P}$ if and only if a_m is real and $\geqq 0$, and $\sum a_m < \infty$; $f(x, y)$ is real if and only if $a_{-m} = a_m$. Thus, by Theorem 3, $\rho(x, y)$ is a Hilbert distance if and only if it is the square root of an expression

$$(14) \quad \sum_{m=1}^{\infty} a_m(2 - e^{im(x-y)} - e^{-i(x-y)}) = \sum_{m=1}^{\infty} a_m(e^{imx} - e^{imy})(e^{-imx} - e^{-imy})$$

where

$$(15) \quad a_m \geqq 0 \quad \text{and} \quad \sum_m a_m < \infty.$$

If we prefer to look upon $\rho(x, y)$ as a function of the one variable $x - y$ it is more appropriate to write (15) in the form

$$(16) \quad 2\sum_{m=1}^{\infty} a_m(1 - \cos m(x - y)) = 4\sum_{m=1}^{\infty} a_m\left(\sin \frac{m}{2}(x - y)\right)^2.$$

This result can be extended to our spaces in general in terms of expansions into "generalized spherical harmonics" as given by E. Cartan and H. Weyl.[4] If $\mathfrak{S}$ is a space satisfying lemma 1 with a transitive group $\mathfrak{C}$ of motions then there exists on $\mathfrak{S}$ a complete orthonormal system of continuous functions of the following description:

(i) corresponding to each k, $k = 0, \pm 1, \pm 2, \pm 3, \cdots$ there exist a finite rectangular system of functions $\varphi_{k,m\mu}(P)$ with

$$(17) \quad \begin{aligned} m &= 1, \cdots, l, \; l = l(k); \\ \mu &= 1, \cdots, h, \; h = h(k), \end{aligned}$$

and a quadratic system of functions $u_{k,m\alpha}(P)$ on $\mathfrak{C}$ with $m, \alpha = 1, \cdots, l = l(k)$, such that

$$(18) \quad \varphi_{k,m\mu}(sP) = \sum_{\alpha=1}^{l} u_{k,m\alpha}(s)\varphi_{k,\alpha\mu}(P).$$

(ii) For each k, $u_{k,m\alpha}(s)$ is an irreducible unitary representation of $\mathfrak{C}$, and for different values of k the representations are inequivalent.

(iii) for $k = 0$, $l = h = 1$, and $\varphi_{0,11}(P) = 1$, $u_{0,11}(s) = 1$; also $l(-k) = l(k)$, $h(-k) = h(k)$, and

$$\varphi_{-k,m\mu}(P) = \overline{\varphi_{k,m\mu}(P)}, \qquad u_{-k,m\alpha}(s) = \overline{u_{k,m\alpha}(s)}.$$

Now the system of functions

$$\varphi_{p,m\mu}(P) \cdot \overline{\varphi_{q,n\nu}(Q)}$$

is an orthonormal system on $\mathfrak{S} \times \mathfrak{S}$, and thus $f(P, Q)$ has an expansion

$$(19) \quad \sum_{p,m,\mu,q,n,\nu} a_{pq,m\mu,n\nu}\varphi_{p,m\mu}(P)\overline{\varphi_{q,n\nu}(Q)}.$$

[4] H. Weyl, *Harmonics on homogeneous manifolds*, Annals of Math., 35 (1934), 486–494.

Replacing P and Q by sP and sQ respectively and using (18) we obtain as a necessary and sufficient condition for group invariance the system of relations

$$a_{pq,\alpha\mu,\beta\nu} = \sum_{m=1}^{l(p)} \sum_{n=1}^{l(q)} a_{pq,m\mu,n\nu}\, u_{p,m\alpha}(s)\overline{u_{q,n\beta}(s)}$$

and this is equivalent with

$$\sum_{n=1}^{l(q)} a_{pq,\alpha\mu,n\nu}\, u_{q,\beta n}(s) = \sum_{m=1}^{l(p)} a_{pq,m\mu,\beta\nu}\, u_{p,m\alpha}(s). \tag{20}$$

Since the functions $\{u(s)\}$ are linearly independent the comparison of coefficients on both sides of (20) will show that $a_{pq,m\mu,n\nu}$ vanishes if $p \neq q$ or if $m \neq n$, and that for $p = q$ and $m = n$ its value is independent of m. This leads to writing (19) in the form

$$\sum_{k=-\infty}^{\infty} \left(\sum_{\mu,\nu=1}^{h(k)} a_{k,\mu\nu}\, \varphi_{k,\mu\nu}(P, Q) \right) \tag{21}$$

where

$$\varphi_{k,\mu\nu}(P, Q) = \sum_{m=1}^{l(k)} \varphi_{k,m\mu}(P)\overline{\varphi_{k,m\nu}(Q)}. \tag{22}$$

We can now express the properties of $\mathfrak{P}$ in terms of the coefficients $a_{k,\mu\nu}$. Property 1) of $\mathfrak{P}$ simply means that for each k, the matrix

$$\| a_{k,\mu\nu} \|_{\mu,\nu=1,\ldots,h(k)} \tag{23}$$

is hermitian. Property 2) is automatically fulfilled, and property 3) means that (23) is non-negative definite. In fact putting in (3)

$$\rho(P) = \sum_{\mu=1}^{h} x_\mu \overline{\varphi_{k,m\mu}(P)}$$

we obtain

$$\sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\, x_\mu \overline{x_\nu} \geqq 0. \tag{24}$$

Conversely if $\rho(P)$ is a finite sum of the form

$$\sum_{p,m,\mu} x_{p,m\mu}\, \overline{\varphi_{p,m\mu}(P)} \tag{25}$$

the left side of (3) is

$$\sum_{p,m} \left(\sum_{\mu,\nu=1}^{h} a_{p,\mu\nu}\, x_{p,m\mu}\, \overline{x_{p,m\nu}} \right)$$

and this is $\geqq 0$ if (24) holds. But finite sums of the form (25) are dense in the family of all function $\rho(P)$ and thus (24) implies property 3). Finally by property (iii) of $\{\varphi(P)\}$, $f(P, Q)$ is real if and only if

$$\overline{a_{k,\nu\mu}} = a_{-k,\mu\nu}. \tag{26}$$

Thus a function $f(P, Q)$ of the real class $\mathfrak{P}$ has an expansion of the form

$$a_0 + \sum_{k=1}^{\infty}\left(\sum_{\mu,\nu=1}^{h(k)} a_{k,\mu\nu}(\varphi_{k,\mu\nu}(P, Q) + \overline{\varphi_{k,\mu\nu}(P, Q)}\right) \tag{27}$$

where each matrix (23) is a semi-definite hermitian matrix. Before proceeding we require the following

LEMMA 2. *There exists an array of real numbers* $r_{n,k}$, $n = 1, 2, \cdots$, $k = 0, \pm 1, \pm 2, \cdots$, *with the following properties*: (i) $0 \leqq r_{n,k} \leqq 1$, (ii) $\lim_{n\to\infty} r_{n,k} = 1$, (iii) *for each* n *only a finite number of coefficients* $r_{n,k}$ *is* $\neq 0$, *and what is decisive*, (iv) *if* (21) *is the expansion of a continuous function* $f(P, Q)$, *then the sequence of functions*

$$f_n(P, Q) = \sum_{k=-\infty}^{\infty} r_{n,k}\left(\sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\varphi_{k,\mu\nu}(P, Q)\right) \tag{28}$$

is uniformly convergent towards $f(P, Q)$ *as* $n \to \infty$.

The proof of the lemma can be carried out along familiar lines and will be omitted.[5]

Now, if (23) is hermitian the number

$$\lambda_k = \sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\varphi_{k,\mu\nu}(\Phi, \Phi)$$

is $\geqq 0$ and

$$\left|\sum_{\mu,\nu=1}^{h} a_{k,\mu\nu}\varphi_{k,\mu\nu}(P, Q)\right| \leqq \lambda_k ;$$

therefore

$$0 \leqq \sum_{k=-\infty}^{\infty} r_{n,k}\lambda_k \leqq f_n(\Phi, \Phi) \leqq M$$

where M is independent of n. Letting $n \to \infty$ we obtain $\sum_k \lambda_k \leqq M$, and we hence conclude that the series (21) and (27) are absolutely and uniformly convergent. Therefore we have for

$$f(\Phi, \Phi) - f(P, Q) \equiv \tfrac{1}{2}f(P, P) + \tfrac{1}{2}f(Q, Q) - f(P, Q)$$

the series

$$\sum_{k=1}^{\infty}\left(\sum_{\mu,\nu=1}^{h(k)} a_{k,\mu\nu}\psi_{k,\mu\nu}(P, Q)\right) \tag{29}$$

where

$$\psi_{k,\mu\nu}(P, Q) = \sum_{m=1}^{l(k)} (\varphi_{k,m\mu}(P) - \varphi_{k,m\mu}(Q))(\overline{\varphi_{k,m\nu}(P)} - \overline{\varphi_{k,m\nu}(Q)}). \tag{30}$$

[5] See S. Bochner and J. v. Neumann, *Almost periodic functions in groups*, Trans. Amer. Math. Soc., 37 (1935), 21–50, esp. Part III; H. Weyl, loc. cit., p. 498–499.

Altogether, on the basis of Theorem 3 the following theorem can now be easily verified.

THEOREM 4. *A non-negative function $\rho(P, Q)$ is a group invariant Hilbert distance if and only if its square is an absolutely and uniformly convergent series of the form* (29) *in which such matrix* (23) *is non-negative hermitian and $\psi_{k,\mu\nu}$ has the value* (30).

We observe that our series (29) is a generalization of series (14).

THEOREM 5. *The Hilbert distance of Theorem 4 is certainly a proper distance if each matrix* (23) *is strictly positive definite.*

PROOF: In fact, if (23) is strictly positive then $\rho(P, Q)$ can only vanish if for each k, m, α

$$\varphi_{k,m\alpha}(P) - \varphi_{k,m\alpha}(Q) = 0.$$

But the latter fact implies that every continuous function on $\mathfrak{S}$ assumes equal values on the points P and Q, and this can only happen if P and Q are identical.

IV. SOME SPECIFIC CASES

Formulas (27) and (29) are greatly simplified if for all values of k we have $h(k) = 1$. In this case we can write

$$f(P, Q) = a_0 + \sum_{k=1}^{\infty} a_k(\varphi_k(P, Q) + \overline{\varphi_k(P, Q)})$$

where

$$\varphi_k(P, Q) + \overline{\varphi_k(P, Q)} = \sum_{m=1}^{l(k)} (\varphi_{k,m}(P)\overline{\varphi_{k,m}(Q)} + \overline{\varphi_{k,m}(P)}\varphi_{k,m}(Q)); \tag{31}$$

also the expression

$$\sum_{k=1}^{\infty} a_k(2\varphi_k(\Phi, \Phi) - \varphi_k(P, Q) - \overline{\varphi_k(P, Q)}) \tag{32}$$

for $\rho(P, Q)^2$ shows a strong resemblance to the left side of (16). The resemblance is most pronounced if $\mathfrak{S}$ is the $(m-1)$-dimensional unit sphere in m-dimensional Euclidean space. In this case (31) is but for a numerical factor depending on k (and m) the expression

$$T_k^{(\lambda)}(\cos\vartheta)$$

where $T_k^{(\lambda)}$ is an ultraspherical polynomial, $\lambda = m - 2/2$, and ϑ is the geodesic distance between the points P, Q. The expression

$$\sum_{k=0}^{\infty} b_k T_k^{(\lambda)}(\cos\vartheta)$$

with $b_k \geqq 0$ and $\sum_k b_k T_k^{(\lambda)}(1) < \infty$, for the most general positive definite func-

tion depending on the geodesic distance alone, has been given previously by Schoenberg.[6]

More special than the assumption $h(k) = 1$ is the assumption $l(k) = 1$. It corresponds to the case of the group $\mathfrak{C}$ being an Abelian group, and in this case we simply have

$$f(P, Q) = \sum_{k=0}^{\infty} a_k(\varphi_k(P)\overline{\varphi_k(Q)} + \overline{\varphi_k(P)}\varphi_k(Q))$$

with $a_k \geqq 0$.

The symbol sP is a real multiplication of P by s if $\mathfrak{S}$ is a group space (in which case we write x, y, etc. instead of P, Q, etc.) and $\mathfrak{C}$ is its group of left transformations $x \to sx$. Our assumptions concerning $\mathfrak{S}$ are now that $\mathfrak{S}$ is a compact separable group, the invariant measure on it being now uniquely determined. In this case, $h(k) = l(k)$ and but for the factor $l(k)^{1/2}$, $\varphi_{k,\mu\nu}(x)$ is $u_{k,\mu\nu}(x)$. Since

$$\sum_{m=1}^{l} u_{k,m\mu}(x)\overline{u_{k,m\nu}(y)} = \sum_{m=1}^{l} u_{k,\nu m}(y^{-1})u_{k,m\mu}(x)$$

we see that, except for the factor $l(k)$,

$$\varphi_{k,\mu\nu}(x, y) = u_{k,\nu\mu}(y^{-1}x),$$

and hence we obtain the following result.

THEOREM 6. *If $\mathfrak{C}$ is a compact separable group, then $f(x, y)$ is a left invariant member of $\mathfrak{P}$ if and only if $f(x, y) = f(y^{-1}x)$ where*

$$f(t) = \sum_{k=-\infty}^{\infty} \sum_{\mu,\nu=1}^{l(k)} b_{k,\mu\nu} u_{k,\mu\nu}(t); \tag{33}$$

the matrix $|\, b_{k,\mu\nu} \,|$ being non-negative hermitian and the series (33) in k being absolutely convergent.

In order to obtain right invariant functions we have to put $t = yx^{-1}$. Our function is invariant on both sides if $f(s^{-1}y^{-1}xs) = f(y^{-1}x)$ that is if $f(t)$ is a class function. For class functions the expansion (33) depends only on the group characters

$$\chi_k(t) = \sum_{u=1}^{l} u_{k,\mu\mu}(t).^{7}$$

Thus we obtain

THEOREM 7. *If $\mathfrak{C}$ is a compact separable group, then $f(x, y)$ is a group invariant member of $\mathfrak{P}$, if and only if*

$$f(x, y) = \sum_{k=-\infty}^{\infty} b_k \chi_k(yx^{-1})$$

---

[6] I. J. Schoenberg, *On positive definite functions on spheres.* Bull. Amer. Math. Soc., 46 (1940), p. 888.

[7] See Bochner-von Neumann, loc. cit.

where $b_k \geqq 0$, *and* $\sum_k b_k \chi_k(1) < \infty$; *and* $\rho(x, y)$ *is a group invariant Hilbert distance if and only if*

$$\rho(x, y)^2 = \sum_{k=1}^{\infty} c_k \left(1 - \frac{\chi_k(yx^{-1}) + \chi_k(xy^{-1})}{2l(k)}\right)$$

with $c_k \geqq 0$; *the function* $\rho(x, y)$ *is certainly a proper distance if all* c_k *are* > 0.

V. Isometric Imbedding of Riemannian Space into Hilbert Space

In the present section we will draw a conclusion from Theorem 5. We assume that our space $\mathfrak{S}$ is a coordinate space of class C_r, $r \geqq 2$, (continuous partial derivatives of order $\leqq r$), or C_∞ (derivatives of every order) or C_ω (analytic coordinates), and we further assume explicitly that the functions $\{\varphi_{k,m\mu}(P)\}$ belong to the same class. By choosing the coefficients of the matrix (23) sufficiently small we can obtain a series (29) whose sum will belong to the same class on $\mathfrak{S} \times \mathfrak{S}$. Now, by a general theorem,[8] such a distance can be generated by a Riemannian metric. Altogether we have the following

Theorem 8. *If* $\mathfrak{S}$ *is a compact differentiable manifold (of arbitrary dimension and) of class* C_r, $r \geqq 2$, *or* C_∞, *or* C_ω, *if* $\mathfrak{C}$ *is a fixed transitive group of homomorphisms on it, and if the corresponding generalized spherical harmonics* $\{\varphi_{k,m\mu}(P)\}$ *belong to the same class, then there exists on* S *a group invariant positive definite Riemannian metric of class* C_{r-2}, *or* C_∞, *or* C_ω *for which the given space can be isometrically imbedded in real Hilbert space.*

Princeton University.

[8] S. Bochner, *Differentiable and Riemann metric*, Duke Math. Jour., 4 (1938), 51–54.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

UNITÄRINVARIANTE HYPERMAXIMALE NORMALE OPERATOREN

VON HIDEGORÔ NAKANO

(Received November 30, 1939)

Unitärinvariante hypermaximale Hermitesche Operatoren im Hilbertschen Raum sind schon in mehreren Formen gegeben.[1] Ich will in dieser Arbeit Unitärinvariante hypermaximale, normale Operatoren im Hilbertschen Raum $\mathfrak{H}$ in einer anderen Form geben.

Massoperatoren $E(Z)$ eines hypermaximalen, normalen Operators bilden einen abgeschlossenen, Abelschen Ring $\mathfrak{R}$ von Projektionsoperatoren, dessen Definition in der früheren Abhandlung[2] gegeben ist. Zuerst definieren wir die Dimension eines abgeschlossenen, Abelschen Rings. Für einen hypermaximalen, normalen Operator N kann man die komplexe Ebene G in eine Folge von Mengen Z_∞, $Z_1, Z_2, \cdots$ derart zerlegen, dass je zwei der $Z_\infty, Z_1, Z_2, \cdots$ keinen gemeinsamen Punkt besitzen, und der aus allen Projektionsoperatoren $E(Z)$ $(Z \subset Z_n)$ gebildete, abgeschlossene, Abelsche Ring $\mathfrak{R}_n$ von der Dimension n ist. Diese Zerlegung ist bis auf eine Nullmenge über $E(Z)$ eindeutig bestimmt. Diese Zerlegung Z_∞, $Z_1, Z_2, \cdots$ nennen wir das Spektralsystem des hypermaximalen, normalen Operators N. Dafür, dass zwei hypermaximale, normale Operatoren N_1 und N_2 unitär äquivalent seien, ist es notwendig und hinreichend, dass die Klassen der Nullmengen und die Spektralsysteme von N_1 und N_2 miteinander übereinstimmen.

1. Dimension von Ring. Abgeschlossene Abelsche Ringe nennen wir kurz *Ringe*. *$\mathfrak{R}$ sei ein Ring, d.h. je zwei Projektionsoperatoren P und Q in $\mathfrak{R}$ sind vertauschbar, $P - PQ$ gehört zu $\mathfrak{R}$, und für jede Folge $P_1, P_2, \cdots$ in $\mathfrak{R}$ gehören $P_1 \dotplus P_2 \dotplus \cdots$ und $P_1 P_2 \cdots$ beide zu $\mathfrak{R}$.*

DEFINITION 1. Ein Projektionsoperator K heisst mit einem Ring $\mathfrak{R}$ *vertauschbar*, wenn jedes P in $\mathfrak{R}$ mit K vertauschbar ist.

DEFINITION 2. Wenn ein Projektionsoperator K mit einem Ring $\mathfrak{R}$ vertauschbar ist, so bilden KP $(P \in \mathfrak{R})$ auch einen Ring, welchen wir den *Nebenring* von $\mathfrak{R}$ nennen.

Da $\mathfrak{H}$ separabel ist, kann man leicht beweisen, dass die Summe $\sum_{P \in \mathfrak{R}} \{P\}$ auch zu $\mathfrak{R}$ gehört.[3] Diese Summe nennen wir den *Maximaloperator* des Rings $\mathfrak{R}$.

---

[1] Vgl. M. H. Stone, *Linear Transformations in Hilbert Space*, New York (1932). F. Wecken, *Unitärinvarianten selbstadjungierter Operatoren*, Math. Annalen, 116 (1939).

[2] H. Nakano, *Über Abelsche Ringe von Projektionsoperatoren*, Proc. of the Phys.-Math. Soc. of Japan, 21 (1939).

[3] $f_1, f_2, \cdots$ sei überall dicht in $\mathfrak{H}$, und $P_{n1}, P_{n2}, \cdots$ sei eine Folge von Projektions-

Indem man den Maximaloperator von $\mathfrak{R}$ mit M bezeichnet, kann man leicht einsehen, dass zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ dann und nur dann übereinstimmen, wenn $KM = LM$ ist, da der Ring $K\mathfrak{R}$ offenbar mit dem Ring $KM\mathfrak{R}$ übereinstimmt.

DEFINITION 3. Zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ heissen *isomorph*, wenn für jeden Projektionsoperator P in $\mathfrak{R}$ $KP = 0$ mit $LP = 0$ gleiche Bedeutung hat.

DEFINITION 4. Zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ heissen zueinander *orthogonal*, wenn für je zwei Projektionsoperatoren P und Q in $\mathfrak{R}$ stets $(KP)(LQ) = 0$ ist.

Man kann leicht einsehen, dass zwei Nebenringe $K\mathfrak{R}$ und $L\mathfrak{R}$ dann und nur dann zueinander orthogonal, sind wenn $(LM)(KM) = 0$ ist.

DEFINITION 5. Wenn n Nebenringe $K_1\mathfrak{R}$, $K_2\mathfrak{R}$, $\cdots$, $K_n\mathfrak{R}$ eines Rings $\mathfrak{R}$ alle mit $\mathfrak{R}$ isomorph, und je zwei von $K_1\mathfrak{R}$, $K_2\mathfrak{R}$, $\cdots$, $K_n\mathfrak{R}$ zueinander orthogonal sind, und noch $n+1$ solche Nebenringe nicht existieren, so heisst n die *Minimaldimension* von $\mathfrak{R}$. Wenn beliebig viele solche Nebenringe existieren, so heisst $\mathfrak{R}$ *von der Dimension* ∞.

DEFINITION 6. $\mathfrak{R}$ sei ein Ring. Wenn für jedes P in $\mathfrak{R}$ die Minimaldimension des Nebenrings $P\mathfrak{R}$ stets n ist, so heisst $\mathfrak{R}$ *gleichmässig dimensional* mit der Dimension n.

Aus den Definitionen folgt sofort der folgende

SATZ 1. *Wenn die Minimaldimension eines Rings $\mathfrak{R}$ grösser als n ist, so ist die Minimaldimension des Nebenrings $P\mathfrak{R}$ für jedes P in $\mathfrak{R}$ auch grösser als n.*

SATZ 2. *Wenn die Minimaldimensionen von endlich oder abzählbar unendlich vielen Nebenringen $L_1\mathfrak{R}$, $L_2\mathfrak{R}$, $\cdots$ alle grösser als n sind, und $L_iL_j = 0$ $(i \neq j)$ gilt, so ist die Minimaldimension des Nebenrings $(L_1 + L_2 + \cdots)\,\mathfrak{R}$ auch grösser als n.*

BEWEIS: Nach Voraussetzung gibt es $n+1$ zueinander orthogonale Nebenringe $K_i^{(1)}L_i\mathfrak{R}$, $K_i^{(2)}L_i\mathfrak{R}$, $\cdots$, $K_i^{(n+1)}L_i\mathfrak{R}$, welche alle mit $L_i\mathfrak{R}$ $(i = 1, 2, \cdots)$ isomorph sind. Dann sind $n+1$ Nebenringe $(L_1K_1^{(1)} + L_2K_2^{(1)} + \cdots)\mathfrak{R}, \cdots,$ $(L_1K_1^{(n+1)} + L_2K_2^{(n+1)} + \cdots)\mathfrak{R}$ offenbar zueinander orthogonal, und sie sind sogar alle mit $(L_1 + L_2 + \cdots)\mathfrak{R}$ isomorph, denn, aus $(L_1K_1^{(j)} + L_2K_2^{(j)} + \cdots)P = 0$ $(P \in \mathfrak{R})$ folgt $L_1K_1^{(j)}P = 0, L_2K_2^{(j)}P = 0, \cdots$, und weiter $L_1P = 0, L_2P = 0, \cdots$, folglich $(L_1 + L_2 + \cdots)P = 0$, und umgekehrt folgt aus $(L_1 + L_2 + \cdots)P = 0$ $(P \in \mathfrak{R})$ offenbar $(L_1K_1^{(j)} + L_2K_2^{(j)} + \cdots)P = 0$. Daher ist die Minimaldimension des Rings $(L_1 + L_2 + \cdots)\mathfrak{R}$ grösser als n.

SATZ 3. *Wenn die Minimaldimensionen von endlich oder abzählbar unendlich vielen Nebenringen $L_1\mathfrak{R}$, $L_2\mathfrak{R}$, $\cdots$ grösser als n sind, und je zwei von $L_1, L_2, \cdots$*

---

operatoren im Ring $\mathfrak{R}$, derart dass

$$\lim_{i\to\infty} \| P_{ni}f_n \| = \text{l.u.b.}_{R \in \mathfrak{R}} \| Rf_n \|$$

ist. Setzt man $P_n = \sum_{i,j \leqq n} P_{ij}$, so ist $P_1 \leqq P_2 \leqq \cdots$, und alle P_n gehören zu $\mathfrak{R}$. Folglich gehört $\lim_{n\to\infty} P_n = M$ auch zu $\mathfrak{R}$, und es gilt $\| Mf_n \| = \text{l.u.b.}_{R \in \mathfrak{R}} \| Rf_n \|$ $(n = 1, 2, \cdots)$. Daher ist M der Maximaloperator in $\mathfrak{R}$.

miteinander vertauschbar sind, so ist die Minimaldimension des Nebenrings $(L_1 \dotplus L_2 \dotplus \cdots)\mathfrak{R}$ *grösser als* n.

BEWEIS: Setzt man

$$\overline{L}_n = L_n - L_n(L_1 \dotplus L_2 \dotplus \cdots \dotplus L_{n-1}),$$

so gilt

$$\overline{L}_i \overline{L}_j = 0 \qquad (i \neq j)$$

$$L_1 \dotplus L_2 \dotplus \cdots = \overline{L}_1 + \overline{L}_2 + \cdots,$$

und nach dem Satz 1 ist die Minimaldimension von $\overline{L}_n$ auch grösser als n. Daher folgt dieser Satz sofort aus dem Satz 3.

SATZ 4. *Wenn ein Ring* $\mathfrak{R}$ *von der Minimaldimension* n *ist, so gibt es zwei Projektionsoperatoren* P *und* Q *in* $\mathfrak{R}$, *derart dass* $PQ = 0$, *der Nebenring* $P\mathfrak{R}$ *gleichmässig dimensional mit den Dimension* n, *die Minimaldimension des Nebenrings* $Q\mathfrak{R}$ *grösser als* n *ist, und* $P + Q$ *mit dem Maximaloperator* M *von* $\mathfrak{R}$ *übereinstimmt.*

BEWEIS: Wenn $\mathfrak{R}$ gleichmässig dimensional mit der Dimension n ist, so braucht man nur $P = M$ zu setzen. Im anderen Falle bilden alle Projektionsoperatoren R in $\mathfrak{R}$, für welche die Minimaldimension des Nebenrings $R\mathfrak{R}$ grösser als n ist, nach den Sätzen 1 und 3 einen Unterring $\mathfrak{R}_1$ von $\mathfrak{R}$. Bezeichnet man den Maximaloperator von $\mathfrak{R}_1$ mit Q, so gehört Q zu $\mathfrak{R}_1$, und die Minimaldimension des Nebenrings $Q\mathfrak{R}$ ist grösser als n, und für $P = M - Q$ ist der Nebenring $P\mathfrak{R}$ offenbar gleichmässig dimensional mit der Dimension n.

2. Unitärinvarianz. Es sei N ein hypermaximaler, normaler Operator mit dem Massoperator $E(Z)$. Wenn man mit $\mathfrak{R}$ den aus allen $E(Z)$ gebildeten Ring bezeichnet, so kann man nach dem Satz 4 die komplexe Ebene G in abzählbar unendlich viele, messbare Punktmengen $Z_\infty, Z_1, Z_2, \cdots$ derart zerlegen, dass der Nebenring $E(Z_i)\mathfrak{R}$ gleichmässig dimensional mit der Dimension $i(= \infty, 1, \cdots)$ ist, und $G = Z_\infty + Z_1 + Z_2 + \cdots$, $Z_iZ_j = 0$ $(i \neq j)$ gilt. Diese Zerlegung $Z_\infty, Z_1, Z_2, \cdots$ der komplexen Ebene G nennen wir das *Spektralsystem* des hypermaximalen, normalen Operators N. Man kann leicht einsehen, dass das Spektralsystem bis auf eine Nullmenge über $E(Z)$ eindeutig bestimmt ist.

SATZ 5. *Dafür, dass zwei hypermaximale normale Operatoren* N_1 *und* N_2 *unitär äquivalent seien, d.h. dafür, dass für einen passenden unitären Operator* U $N_1 = U^*N_2U$ *bestehe, ist es notwendig und hinreichend, dass die Klassen der Nullmengen und die Spektralsysteme von* N_1 *und* N_2 *bis auf eine Nullmenge übereinstimmen.*

Mit $E_1(Z)$ und $E_2(Z)$ bezeichnen wir den Massoperator bzw. von N_1 und N_2. Wenn für einen unitären Operator U $N_1 = U^*N_2U$ ist, so erhält man nach der Eigenwertdarstellung

$$\int_G z\, d(E_1(Z)f, g) = \int_G z\, d(U^*E_2(Z)Uf, g)$$

für jedes f im Definitionsbereich von N_1 und jedes g in $\mathfrak{H}$, folglich nach der Eindeutigkeit des Massoperators

$$E_1(Z) = U^* E_2(Z) U.$$

Daher kann man leicht einsehen, dass die Klassen der Nullmengen über $E_1(Z)$ und $E_2(Z)$ und die Spektralsysteme von N_1 und N_2 bis auf eine Nullmenge übereinstimmen.

Für den Beweis der Hinreichendheit brauchen wir folgende Hilfssätze.

HILFSSATZ 1. *$\mathfrak{R}$ sei ein Ring mit dem Maximaloperator M. Es gibt mindestens ein Element f in $\mathfrak{H}$ derart, dass für jedes $R(\neq 0)$ in $\mathfrak{R}$ stets $Rf \neq 0$ ist. Ein solches Element f nennen wir ein eigentliches Element in bezug auf $\mathfrak{R}$.*

BEWEIS: Da $\mathfrak{H}$ separabel ist, gibt es eine Folge von Elementen $f_1, f_2, \cdots$ derart, dass $\mathfrak{H}$ von $f_1, f_2, \cdots$ aufgespannt wird, und $\|f_1\|^2 + \|f_2\|^2 + \cdots$ konvergiert. Alle Projektionsoperatoren R in $\mathfrak{R}$, für welche $Rf_n = 0$ ist, bilden zusammen einen Unterring $\mathfrak{R}_n$ von $\mathfrak{R}$. Wenn man den Maximaloperator von $\mathfrak{R}_n$ mit E_n bezeichnet, so kann man leicht einsehen, dass $\mathfrak{R}_n$ mit dem Nebenring $E_n\mathfrak{R}$ übereinstimmt. Daher ist $(M - E_n)Rf_n \neq 0$ für jedes R in $\mathfrak{R}$ wenn $(M - E_n)R \neq 0$ ist. Setzt man $E'_n = E_1 E_2 \cdots E_n$, so konvergiert

$$(M - E'_1)f_1 + (E'_1 - E'_2)f_2 + (E'_2 - E'_3)f_3 + \cdots$$

nach einem Element f_0, da $\|(E'_{n-1} - E'_n)f_3\|^2 \leqq \|f_3\|^2$ ist, und $\|f_1\|^2 + \|f_2\|^2 + \cdots$ konvergiert. Dieses Element f_0 ist ein eigentliches Element in bezug auf $\mathfrak{R}$, denn, wenn für einen R in $\mathfrak{R}$ $Rf_0 = 0$ gilt, dann ist zugleich

$$(M - E'_1)Rf_0 = 0, \qquad (E'_1 - E'_2)Rf_0 = 0, \qquad (E'_2 - E'_3)Rf_0 = 0, \cdots,$$

folglich muss

$$(M - E'_1)R = 0, \quad (M - E_2)(E'_1 - E'_2)R = 0, \quad (M - E_3)(E'_2 - E'_3)R = 0, \cdots$$

sein. Da

$$\begin{aligned}(M - E_n)(E'_{n-1} - E'_n) &= (M - E_n)(E_1 \cdots E_{n-1} - E_1 \cdots E_{n-1}E_n) \\ &= E_1 \cdots E_{n-1} - E_1 \cdots E_{n-1}E_n = E'_{n-1} - E'_n\end{aligned}$$

ist, erhält man

$$(M - E'_n)R = 0, \qquad \text{d.h.} \qquad MR = E'_n R$$

folglich

$$MR = R \lim_{n\to\infty} E'_n.$$

Andererseits gilt $(\lim_{n\to\infty} E'_n)f_i = 0$ $(i = 1, 2, \cdots)$, und $f_1, f_2, f_3, \cdots$ spannen $\mathfrak{H}$ auf. Daher muss $\lim_{n\to\infty} E'_n = 0$ sein, folglich $R = MR = R(\lim_{n\to\infty} E'_n) = 0$.

DEFINITION 7. Zwei in bezug auf $\mathfrak{R}$ eigentliche Elemente f und g heissen *eigentlich orthgonal*, wenn für jedes R in $\mathfrak{R}$ $(Rf, g) = (f, Rg) = 0$ gilt.

DEFINITION 8. $\mathfrak{R}$ sei ein Ring mit dem Maximaloperator M. Wenn für n eigentliche Elemente $f_1, f_2, \cdots, f_n$ $(n = \infty, 1, 2, \cdots)$, von denen je zwei zueinander eigentlich orthogonal sind, die abgeschlossene lineare Mannigfaltigkeit des Maximaloperators aus allen $\{Rf_1, Rf_2, \cdots, Rf_n\}$ $(R \in \mathfrak{R})$ aufgespannt wird, so heisst $f_1, f_2, \cdots, f_n$ ein *Basensystem* des Rings $\mathfrak{R}$.

HILFSSATZ 2. *Wenn ein Ring $\mathfrak{R}$ mit dem Maximaloperator M gleichmässig dimensional mit endlicher Dimension n ist, so besitzt $\mathfrak{R}$ ein Basensystem, das aus n eigentlichen Elementen besteht.*

BEWEIS: Da die Dimension von $\mathfrak{R}$ n ist, so gibt es zueinander orthogonale, mit $\mathfrak{R}$ isomorphe Nebenringe $K_1\mathfrak{R}, K_2R, \cdots, K_n\mathfrak{R}$ mit den Maximaloperatoren $K_1, K_2, \cdots, K_n$. Wenn man mit f_i ein eigentliches Element in bezug auf $K_i\mathfrak{R}$ bezeichnet, so sind je zwei der $K_1f_1, K_2f_2, \cdots, K_nf_n$ offenbar eigentlich orthogonal. und K_if_i ist ein eigentliches Element in bezug auf $K_i\mathfrak{R}$. Bezeichnet man mit $\bar{K}_i$ den Projektionsoperator der von allen $\{RK_if_i\}$ $(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit, so ist offenbar $M \geqq \bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n$. Wenn $M - (\bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n) \neq 0$ ist, so bilden alle Projektionsoperatoren R in $\mathfrak{R}$, für die $[M - (\bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n)]R = 0$ ist einen Ring mit dem Maximaloperator R_0, und es ist $M - R_0 \neq 0$. Dann ist die Minimaldimension von $(M - R_0)\mathfrak{R}$ grösser als n, was unmöglich ist, da $\mathfrak{R}$ gleichmässig dimensional mit der Dimension n ist. Daher ist $M = \bar{K}_1 + \bar{K}_2 + \cdots + \bar{K}_n$ d.h. $f_1, f_2, \cdots, f_n$ ist ein Basensystem von $\mathfrak{R}$.

HILFSSATZ 3. *Wenn ein Ring $\mathfrak{R}$ mit dem Maximaloperator M ein aus n Elementen bestehendes Basensystem $f_1, f_2, \cdots, f_n$ besitzt, so ist $\mathfrak{R}$ von der Minimaldimension n.*

BEWEIS: Wenn man mit $\mathfrak{M}$ die dem Projektionsoperator $\mathfrak{M}$ entsprechende, abgeschlossene, lineare Mannigfaltigkeit bezeichnet, so gibt es einen hypermaximalen, normalen Operator N in $\mathfrak{M}$, dessen Eigenring mit $\mathfrak{R}$ übereinstimmt.[4] Für jedes f_0 in $\mathfrak{M}$ lässt sich jedes Element in der aus allen $Rf_0 (R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltikeit in der Form $\Phi(N)f_0$ darstellen, wo $\Phi(N)$ eine Funktion von N, und $\Phi(N)f_0$ sinnvoll ist.[5] Da $Mf_1, Mf_2, \cdots, Mf_n$ auch ein Basensystem von $\mathfrak{R}$ ist, so kann man annehmen, dass $f_1, f_2, \cdots, f_n$ alle zu $\mathfrak{M}$ gehören, und wir betrachten daher nur den Raum $\mathfrak{M}$. Wenn die

[4] Satz 8 in der Abhandlung: H. Nakano, *Über Abelsche Ringe von Projektionsoperatoren*, Proc. of the Phys.-Math. Soc. of Japan 21 (1939).

[5] Vgl. M. H. Stone, *Linear Transformations in Hilbert Space*, New York (1932). Man kann es auch wie folgt einsehen: Alle über $E(Z)$ messbare Funktionen $\Phi(z)$, für die das Integral $\int_G |\Phi(z)|^2 d\,||E(Z)f_0||^2$ konvergiert, bilden nach dem Riesz-Fisherschen Satz einen Hilbertschen Raum mit dem inneren Produkt $\int_G \Phi_1(z)\overline{\Phi_1(z)}\, d\,||E(Z)f_0||^2$. Dieser Hilbertsche Raum ist mit der aus allen $\Phi(N)f_0$ bestehenden, linearen Mannigfaltigkeit isometrisch, da $(\Phi_1(N)f_0, \Phi_2(N)f_0) = \int_G \Phi_1(z)\overline{\Phi_2(z)}\, d\,||E(Z)f_0||^2$ ist. Daher bilden $\Phi(N)f_0$ eine abgeschlossene lineare Mannigfaltigkeit, die offenbar von allen $E(Z)f_0$ aufgespannt wird.

Minimaldimension von $\mathfrak{R}$ grösser als n ist, so kann man wie im obigen Beweis leicht einsehen, dass zueinander eigentlich orthogonale Elemente $\varphi_1, \varphi_2, \cdots, \varphi_m$ $(m > n)$ existieren. Da $\mathfrak{M}$ von allen $\{Rf_1, Rf_2, \cdots, Rf_n\}(R \in \mathfrak{R})$ aufgespannt wird, so kann man schreiben:

$$(*) \qquad \begin{aligned} \varphi_1 &= \Phi_{11}f_1 + \cdots + \Phi_{1n}f_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \varphi_m &= \Phi_{m1}f_1 + \cdots + \Phi_{mn}f_n \end{aligned} \qquad (m > n)$$

wo Φ_{ij} passende Funktionen von N sind, daher sind Φ_{ij} alle hypermaximal, ihr Eigenring ist in $\mathfrak{R}$ enthalten, und je zwei der $\Phi_{ij}\begin{pmatrix} i = 1, 2, \cdots, m \\ j = 1, 2, \cdots, n \end{pmatrix}$ sind miteinander vertauschbar. Da Φ_{ij} hypermaximal ist, so gibt es für jede positive Zahl p einen Projektionsoperator $R_{ij}^{(p)}$ in $\mathfrak{R}$ derart, dass $\| \Phi_{ij}R_{ij}^{(p)}f \| \leqq p \| R_{ij}^{(p)}f \|$ für jedes f in $\mathfrak{M}$, und $\lim_{p\to\infty} R_{ij}^{(p)} = M$ ist. Setzt man $R^{(p)} = \prod_{ij} R_{ij}^{(p)}$, so gilt $\lim_{p\to\infty} R^{(p)} = M$. Daher ist $R^{(p)} \neq 0$ für eine genügend grosse, positive Zahl p, und $\Phi_{ij}R^{(p)}$ sind alle beschränkt. Aus (*) erhält man

$$\begin{aligned} R^{(p)}\varphi_1 &= \Phi_{11}R^{(p)}f_1 + \cdots + \Phi_{1n}R^{(p)}f_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ R^{(p)}\varphi_m &= \Phi_{m1}R^{(p)}f_1 + \cdots + \Phi_{mn}R^{(p)}f_n \end{aligned} \qquad (m > n)$$

wobei $R^{(p)}\varphi_i \neq 0$ $(i = 1, 2, \cdots, m)$ ist, da alle $\varphi_1, \varphi_2, \cdots, \varphi_m$ eigentliche Elemente in bezug auf $\mathfrak{R}$ sind. Daher verschwinden nicht alle Operatoren $\Phi_{ij}R^{(p)}$ identisch in $\mathfrak{M}$. Nun kann man wie in der Algebra beschränkte Operatoren $\Psi_1, \Psi_2, \cdots, \Psi_m$ in der Form von Determinanten mit den Elementen $\Phi_{ij}R^{(p)}$ derart bestimmen, dass

$$\Psi_1R^{(p)}\varphi_1 + \Psi_2R^{(p)}\varphi_2 + \cdots + \Psi_mR^{(p)}\varphi_m = 0$$

ist, und nicht alle $\Psi_1R^{(p)}, \Psi_2R^{(p)}, \cdots, \Psi_mR^{(p)}$ identisch in $\mathfrak{M}$ verschwinden. Da je zwei der $\varphi_1, \varphi_2, \cdots, \varphi_m$ zueinander eigentlich orthogonal sind, so kann man durch Eigenwertdarstellung leicht einsehen, dass je zwei der $\Psi_1R^{(p)}\varphi_1, \Psi_2R^{(p)}\varphi_2, \cdots, \Psi_mR^{(p)}\varphi_m$ zueinander orthogonal sind. Daher muss

$$\Psi_1R^{(p)}\varphi_1 = 0, \qquad \Psi_2R^{(p)}\varphi_2 = 0, \cdots, \Psi_mR^{(p)}\varphi_m = 0$$

sein. Da nicht alle Operatoren $\Psi_1R^{(p)}, \Psi_2R^{(p)}, \cdots, \Psi_mR^{(p)}$ identisch verschwinden, kann man annehmen, dass $\Psi_1R^{(p)}$ nicht identisch verschwindet. Dann gibt es einen Projektionsoperator R_1 $(\neq 0)$ in $\mathfrak{R}$, sodass für jedes f in $\mathfrak{M}$ aus $R_1f \neq 0$, $\Psi_1R^{(p)}Rf \neq 0$ folgt, da $\Psi_1R^{(p)}$ auch eine Funktion von N ist. Für dieses R_1 gilt $\Psi_1R^{(p)}R_1\varphi_1 = 0$, folglich $R_1\varphi_1 = 0$, was unmöglich ist, weil φ_1 ein eigentliches Element in bezug auf $\mathfrak{R}$ ist.

HILFSSATZ 4. *Ein Ring $\mathfrak{R}$ mit der Dimension ∞ hat ein Basensystem $f_1, f_2, \cdots$.*

BEWEIS: $f_1, f_2, \cdots$ seien in $\mathfrak{H}$ überall dicht. Alle Projektionsoperatoren R in $\mathfrak{R}$, für welche $Rf_1 = 0$ ist, bilden einen Ring mit dem Maximaloperator R_1.

Es gibt nach dem Hilfssatz 2 ein eigentliches Element φ_1 in bezug auf dem Nebenring $R_1\mathfrak{R}$. $g_1 = f_1 + R_1\varphi_1$ ist dann ein eigentliches Element in bezug auf $\mathfrak{R}$, denn, da $(Rf_1, RR_1\varphi_1) = (RR_1f_1, \varphi_1) = 0$ ist, folgt aus $R(f_1 + R_1\varphi_1) = 0$ für ein R in $\mathfrak{R}Rf_1 = 0, RR_1\varphi_1 = 0$, und weiter $RR_1 = R, RR_1 = 0$, d. h. $R = 0$.

Bezeichnet man mit K_1 den Projektionsoperator der von allen $\{R(f_1 + R_1\varphi_1)\}$ $(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit, so ist der Nebenring $(1 - K_1)\mathfrak{R}$ mit $\mathfrak{R}$ isomorph, denn, wäre $(1 - K_1)R = 0$ für ein $R(\neq 0)$ in $\mathfrak{R}$, so müsste die Minimaldimension des Nebenrings $K_1R\mathfrak{R}$ gleich 1 sein, da $f_1 + R_1\varphi_1$ ein Basensystem von $K_1R\mathfrak{R}$ ist, was nach dem Satz 1 unmöglich ist.

Alle Projektionsoperatoren R in $\mathfrak{R}$, für welche $R(1 - K_1)f_2 = 0$ ist, bilden einen Ring mit dem Maximaloperator R_2. Es gibt nach dem Hilfssatz 2 ein eigentliches Element φ_2 in bezug auf dem Nebenring $R_2\mathfrak{R}$. $g_2 = (1 - K_1)(f_2 + R_2\varphi_2)$ ist dann ein eigentliches Element in bezug auf $\mathfrak{R}$, das zu $f_1 + R_1\varphi_1$ eigentlich orthogonal ist.

Bezeichnet man mit K_2 den Projektionsoperator der aus allen $\{R(1 - K_1)g_2\}$ $(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit, so kann man auch wie oben schliessen, dass $[1 - (K_1 + K_2)]\mathfrak{R}$ mit $\mathfrak{R}$ isomorph ist. Ähnlich erhält man eine Folge von eigentlichen Elementen $g_1, g_2, \cdots$, von denen je zwei zueinander eigentlich orthogonal sind. Bezeichnet man mit M den Maximaloperator von $\mathfrak{R}$, so gehören alle $Mf_1, Mf_2, \cdots, Mf_n$ zu der von allen $\{Rg_1, Rg_2, \cdots, Rg_n\}(R \in \mathfrak{R})$ aufgespannten, abgeschlossenen, linearen Mannigfaltigkeit. Daher ist $g_1, g_2, \cdots$ ein Basensystem des Rings $\mathfrak{R}$.

Beweis des Satzes 5. N_1 und N_2 seien zwei hypermaximale, normale Operatoren mit den Massoperatoren. $E_1(Z)$ und $E_2(Z)$. Jede Nullmenge über einem der $E_i(Z)$ sei auch Nullmenge über dem anderen, und Z_∞, $Z_1, Z_2, \cdots$ sei ein gemeinsames Spektralsystem von N_1 und N_2. Wenn man die Eigenringe von N_1 und N_2 bzw. mit $\mathfrak{R}_1$ und $\mathfrak{R}_2$ bezeichnet, so besteht der Nebenring $E_1(Z_n)\mathfrak{R}_1$ aus allen $\{E_1(Z)\}(Z \subset Z_n)$, und der Nebenring $E_2(Z_n)\mathfrak{R}_n$ aus allen $\{E_2(Z)\}(Z \subset Z_n)(n = \infty, 1, 2, \cdots)$. $f_{n,1}, f_{n,2}, \cdots, f_{n,n}$ und $g_{n,1}$, $g_{n,2}, \cdots, g_{n,n}$ seien Basensysteme bzw. von $E_1(Z_n)\mathfrak{R}_1$ und $E_2(Z_n)\mathfrak{R}_2$ $(n = \infty, 1, 2, \cdots)$. Für beliebige beschränkte Funktionen

$$\Phi_{ij}(N_i) \begin{pmatrix} i = \infty, 1, 2, \cdots \\ j = 1, 2, 3, \cdots \end{pmatrix} \quad i \geqq j$$

sind dann je zwei der $\Phi_{ij}(N_1)E_i(Z_i)f_{i,j}$ zueinander orthogonal. Folglich bilden alle endlichen Summen aus ihnen eine lineare Mannigfaltigkeit $\mathfrak{R}_1$. $\mathfrak{R}_1$ ist überall dicht in $\mathfrak{H}$, da Massoperator $E_1(Z)$ auch eine beschränkte Funktion von N_1 ist. Die zwei Mengenfunktionen $\| E_1(Z)E_1(Z_n)f_{n,j} \|^2$ und $\| E_2(Z)E_2(Z_n)g_{n,j} \|^2$ verschwinden für eine Menge Z beide zugleich, wenn eine von ihnen verschwindet. Daher gibt es eine messbare Funktion $\Psi_{nj}(z)$, so dass

$$\| E_1(Z)E_1(Z_n)f_{n,j} \|^2 = \int_G | \Psi_{nj}(z) |^2 d \| E_2(Z)E_2(Z_n)g_{n,j} \|^2$$

$$\begin{pmatrix} n = \infty, 1, 2, \cdots \\ j = 1, 2, 3, \cdots \end{pmatrix} \quad n \geqq j$$

ist.[6] Dann ist $f'_{n,j} = \Psi_{n,j}(N_2)E_2(Z_n)g_{n,j}$ $(j = 1, 2, \cdots, n)$ auch ein Basensystem von $E_2(Z_n)\mathfrak{R}_2$, und es gilt

$$|| E_1(Z)E_1(Z_n)f_{n,j} || = || E_2(Z)E_2(Z_n)f'_{n,j} ||.$$

Wenn man jeder beliebigen, endlichen Summe $f_1 = \sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j}$ die entsprechende Summe $f_2 = \sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j}$ zuordnet, so ist diese Zuordnung längentreu, da

$$\begin{aligned} ||f_1||^2 &= || \sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j} ||^2 = \sum_{n,j} || \Phi_{nj}(N_1)E_1(Z_n)f_{n,j} ||^2 \\ &= \sum_{n,j} \int_{Z_n} | \Phi_{nj}(z) |^2 d || E_1(Z)E_1(Z_n)f_{n,j} ||^2 \\ &= \sum_{n,j} \int_{Z_n} | \Phi_{nj}(z) |^2 d || E_2(Z)E_2(Z_n)f'_{n,j} ||^2 \\ &= || \sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j} ||^2 = ||f_2||^2. \end{aligned}$$

Die $\sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j}$ sind überall dicht in $\mathfrak{H}$. Daher kann man diese Zuordnung zu einer unitären Zuordnung $f_2 = Uf_1$ eindentig erweitern. Für jede charakteristische Funktion $\Lambda(z)$ einer messbaren Menge Z ist $E_1(Z) = \Lambda(N_1)$, $E_2(Z) = \Lambda(N_2)$. Daher erhält man

$$\begin{aligned} UE_1(Z)\{\sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j}\} &= U\{\sum_{n,j} \Phi_{nj}(N_1)\Lambda(N_1)E_1(Z_n)f_{n,j}\} \\ &= \sum_{n,j} \Phi_{nj}(N_2)\Lambda(N_2)E_2(Z_n)f'_{n,j} = E_2(Z)\{\sum_{n,j} \Phi_{nj}(N_2)E_2(Z_n)f'_{n,j}\} \\ &= E_2(Z)U\{\sum_{n,j} \Phi_{nj}(N_1)E_1(Z_n)f_{n,j}\}. \end{aligned}$$

Folglich gilt für jede messbare Menge Z

$$E_1(Z) = U^*E_2(Z)U,$$

und hieraus kann man durch Eigenwertdarstellung leicht schliessen, dass $N_1 = U^*N_2U$ ist.

IMPERIAL UNIVERSITY, TOKYO, JAPAN.

[6] Radon-Nykodymscher Satz. Vgl. S. Saks, *Theory of the Integral*, Warszawa (1937), S. 36.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ÜBER DEN BEWEIS DES STONESCHEN SATZES

VON HIDEGORÔ NAKANO

(Received February 13, 1940)

Der wohlbekannte Stonesche[1]

SATZ. *$U_t(-\infty < t < +\infty)$ sei eine Schar unitärer Operatoren im Hilbertschen Raum $\mathfrak{H}$. Wenn stets*

$$U_t U_s = U_{t+s}$$

gilt, und für jedes f und g in $\mathfrak{H}$ $(U_t f, g)$ als eine Funktion von t messbar ist, so gibt es eine Zerlegung der Einheit $E(\lambda)$, derart dass

$$U_t = \int_{-\infty}^{+\infty} e^{it\lambda}\, dE(\lambda).$$

wurde von J. von Neumann,[2] M. H. Stone,[3] S. Bochner,[4] F. Riesz[5] und B. von Sz. Nagy[6] bewiesen. Unter diesen Beweisen beruht nur der von Neumannsche auf keinem Satz über Fouriersche Integrale oder Fouriersche Reihen. In dieser Abhandlung soll der von Neumannsche Beweis wesentlich vereinfacht werden—insbesondere wird ein besonderer Beweis der Stetigkeit von U_t in t und die damit zusammenhängende Bezugnahme auf die reelle Funktionentheorie vermieden.

Beweis. Wie von Neumann Zeigt, kann man durch

$$A = \int_0^{+\infty} e^{-t} U_t\, dt \tag{1}$$

einen beschränkten Operator erhalten. Es gilt dann weiter

$$A^* = \int_{-\infty}^{0} e^{t} U_t\, dt \tag{2}$$

$$U_s A = A U_s = e^s \int_s^{+\infty} e^{-t} U_t\, dt \tag{3}$$

$$AA^* = A^*A, \qquad A + A^* = 2AA^*. \tag{4}$$

Ferner gilt:

$$\text{aus } Af = 0 \text{ folgt } f = 0. \tag{5}$$

[1] M. H. Stone, Proc. Nat. Ac. 16, February 1930. S. 173–174.
[2] J. von Neumann, Annals of Math. 33, 1932. S. 567–573.
[3] M. H. Stone, Annals of Math. 33, 1932. S. 643–648.
[4] S. Bochner, Sitz. Preusz. Akad. 1933.
[5] F. Riesz, Acta Sci. Math. Szeged, 6, 1934.
[6] B. von Sz. Nagy, Math. Annalen 112, 1936. S. 291–294.

Denn wegen (3) folgt für jedes g

$$\int_s^{+\infty} e^{-t}(U_t f, g)\, dt = 0.$$

Da s beliebig ist, gilt $(U_t f, g) = 0$ bis auf eine Nullmenge in t. Dann kann man nach Separabilität von $\mathfrak{H}$ leicht beweisen, dass bis auf eine Nullmenge in t für alle g gleichzeitig $(U_t f, g) = 0$ d.h. $U_t f = 0$, folglich $\|f\| = \|U_t f\| = 0$.

Sei nun

$$W = -2A + 1. \tag{6}$$

Dann folgt aus (4) $W^*W = WW^* = 1$, d.h. W ist unitär. Und wegen (5) folgt aus $Wf = f, f = 0$. Daher ist W nach J. von Neumann,[7] S. 91, 92 die Cayleysche Transformierte eines hypermaximalen Hermiteschen Operators

$$H = \int_{-\infty}^{+\infty} \lambda\, dE(\lambda),$$

für eine geeignete Zerlegung der Einheit $E(\lambda)$. Die ebendort durchgeführten Rechnungen ergeben weiterhin

$$W = \int_{-\infty}^{+\infty} \frac{\lambda - i}{\lambda + i}\, dE(\lambda).$$

Im Hinblick auf (6) folgt hieraus

$$A = \int_{-\infty}^{+\infty} \frac{1}{1 - i\lambda}\, dE(\lambda). \tag{7}$$

Setzt man

$$V_t = \int_{-\infty}^{+\infty} e^{i\lambda t}\, dE(\lambda)$$

für diese Zerlegung der Einheit $E(\lambda)$, so ist V_t unitär und stetig in t, und es gilt

$$V_{t+s} = V_t V_s \tag{8}$$

$$A = \int_0^{+\infty} e^{-t} V_t\, dt \tag{9}$$

$$A^* = \int_{-\infty}^{0} e^{t} V_t\, dt \tag{10}$$

$$V_s A = A V_s = e^s \int_s^{+\infty} e^{-t} V_t\, dt. \tag{11}$$

Setzt man für eine positive Zahl $n(>1)$ $E_n = E(n) - E(-n)$, und bezeichnet den Wertvorrat von E_n mit $\mathfrak{M}_n$, so gilt für jedes f in $\mathfrak{M}_n$ nach (7)

$$A = \int_{-n}^{n} \frac{1}{1 - i\lambda}\, dE(\lambda).$$

[7] J. von Neumann, Math. Annalen 102, 1929. S. 49–131.

Infolgedessen definiert in $\mathfrak{M}_n$

$$A^{-1} = \int_{-n}^{n} (1 - i\lambda)\, dE(\lambda)$$

eine Reciproke A^{-1} von A, die ausserdem beschränkt ist: Für alle f in $\mathfrak{M}_n$ ist

$$\| A^{-1} f \| \leqq 2n \| f \|, \qquad \| Af \| \geqq \frac{1}{2n} \| f \|.$$

(1), (3), (9), (11) verbindend kann man auch schreiben, dass

$$AU_s = e^s A - e^s \int_0^s e^{-t} U_t\, dt$$

$$AV_s = e^s A - e^s \int_0^s e^{-t} V_t\, dt,$$

daher

$$A(U_s - V_s) = -e^s \int_0^s e^{-t}(U_t - V_t)\, dt.$$

Da die abgeschlossene lineare Mannigfaltigkeit $\mathfrak{M}_n$ beide unitäre Operatoren U_s und V_s reduziert, so gilt für jede positive Zahl s

$$\frac{1}{2n} \| (U_s - V_s) f \| \leqq \| A(U_s - V_s) f \| \leqq e^s \int_0^s e^{-t} \| (U_t - V_t) f \|\, dt.^8$$

Setzt man $\varphi(s) = e^{-s} \| (U_s - V_s) f \|$, so ist

$$0 \leqq \frac{1}{2n} \varphi(s) \leqq \int_0^s \varphi(t)\, dt, \tag{12}$$

folglich

$$\frac{1}{2n} \int_0^s e^{-2ns} \varphi(s)\, ds \leqq \int_0^s \left\{ e^{-2ns} \left[\int_0^s \varphi(t)\, dt \right] \right\} ds$$

$$= -\frac{1}{2n} e^{-2ns} \int_0^s \varphi(t)\, dt + \frac{1}{2n} \int_0^s e^{-2ns} \varphi(s)\, ds,$$

und somit

$$\frac{1}{2n} e^{-2ns} \int_0^s \varphi(t)\, dt \leqq 0.$$

Daher muss nach (12) $\varphi(s) = 0$ sein, d.h. $U_s f = V_s f$ für jedes f in $\mathfrak{M}_n$. Für beliebiges f in $\mathfrak{H}$ gilt auch $U_s E_n f = V_s E_n f$. Da $\lim_{n \to \infty} E_n = 1$ ist, erhält man $U_s f = V_s f$, was zu beweisen war.

Hier danke ich Herrn Prof. J. von Neumann für freundliche Verbesserungen und Bemerkungen.

---

[8] Wenn $(Af, g) = \int (W_t f, g)\, dt$ ist, so gilt

$$|(Af, g)| \leqq \left| \int |(W_t f, g)|\, dt \right| \leqq \left| \int \| W_t f \| \cdot \| g \|\, dt \right| \leqq \| g \| \left| \int \| W_t f \|\, dt \right|.$$

Setzt man $g = Af$, so erhält man

$$\| Af \| \leqq \left| \int \| W_t f \|\, dt \right|.$$

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON THE TOPOLOGICAL STRUCTURE OF SOLVABLE GROUPS

BY CLAUDE CHEVALLEY

(Received November 1, 1940)

E. Cartan has shown that a simply connected solvable Lie group is homeomorphic to some cartesian space.[1] We want to investigate the structure of a solvable Lie group which is not necessarily simply connected.

From a well known theorem,[2] it follows that such a group may be considered as the factor group of a solvable simply connected group G by a discrete subgroup D of the center of G. Therefore what we have to do is to look for the possible "situation" of such a sub-group D in the group G.

I. NOTATIONS. GENERALITIES

G being a Lie group, to every element L of its Lie algebra $\mathfrak{L}$ there corresponds a one parametric sub-group $\sigma = \sigma(t)$ of G: we have $\sigma(t_1 + t_2) = \sigma(t_1)\sigma(t_2)$, and L is the tangent vector at the unit element to the differentiable curve described by $\sigma(t)$. Moreover, replacing L by λL, where λ is a real number, does not change the geometric curve $\sigma(t)$, but replaces the parameter t by $\frac{1}{\lambda}t$.

Extending to general Lie groups a notion which is well known for matrices, we shall denote by $\exp L$ the point of parameter 1 on the curve $\sigma(t)$. Therefore we have $(\exp tL)\ (\exp t'L) = \exp\ (t + t')L$.

If we transform the elements of the group $\exp Lt$ by a fixed element σ, we obtain a new one-parametric group $\exp L't$; the mapping $L \to L'$ is a linear mapping A_σ of the Lie algebra $\mathfrak{L}$ into itself; the mapping $\sigma \to \mathsf{A}_\sigma$ is a linear representation of the group G, which is called the *adjoint representation*. Any central element of G is obviously represented by the identity operation.

Suppose that σ describes itself a new one parametric group $\exp Mu$. Then L' becomes a function of u, and we have $L' = L + u[M, L] + \cdots$, the terms not written being of order > 1.

The mapping $L \to [M, L]$, for M fixed, is a linear operation A_M in $\mathfrak{L}$; the mapping $M \to \mathsf{A}_M$ is called the adjoint representation[3] of the Lie algebra; we have

$$\mathsf{A}_{[M,M']} = \mathsf{A}_{M'}\mathsf{A}_M - \mathsf{A}_M\mathsf{A}_{M'}.$$

[1] It follows at once from the lemma proved by *E. Cartan* on page 20 of the pamphlet "La topologie des groupes de Lie" (Exposes de geometrie, Act. scient. et indus. no. 358, Hermann et Cie, Paris, 1936). See also Theorem 83, p. 277 in *Pontrjagin*, "Topological groups" (Princeton University Press, 1939), Theorem 61, p. 225.

[2] See *Pontrjagin*, l. c. 1), Theorem 83, p. 277.

[3] Our notations do not coincide entirely with the usual ones. It is the mapping $M \to -A_M$ which is generally called the adjoint representation.

Let H be a closed invariant sub-group of G; it is also a Lie group, and its Lie algebra $\mathfrak{H}$ is an invariant sub-algebra—or ideal—in $\mathfrak{L}$. Moreover G/H is also a Lie group, whose Lie algebra is the quotient algebra $\mathfrak{L}/\mathfrak{H}$. If L, L' are two elements of $\mathfrak{L}$, and if L is a congruent to L' modulo $\mathfrak{H}$, i.e. if $L' - L \in \mathfrak{H}$ then the elements $\exp L$, $\exp L'$ belong to the same coset of G modulo H.

II. Solvable Groups

A Lie algebra $\mathfrak{L}$ is *solvable*[4] when there exists a chain $\mathfrak{H}_0 \subset \mathfrak{H}_1 \subset \cdots \subset \mathfrak{H}_r$ of ideals in $\mathfrak{L}$, beginning with $\mathfrak{H}_0 = \{0\}$, ending with $\mathfrak{H}_r = \mathfrak{L}$, such that the factor algebras $\mathfrak{H}_i/\mathfrak{H}_{i-1}$ $(1 \leqq i \leqq r)$ are all abelian. Every sub-algebra or factor-algebra of a solvable Lie algebra is again solvable.

The derived algebra $\mathfrak{L}'$ of a solvable Lie algebra $\mathfrak{L} = \mathfrak{L}_n$ of dimension $n > 0$ is always different from $\mathfrak{L}$. Since $\mathfrak{L}/\mathfrak{L}'$ is abelian, every linear sub-space of $\mathfrak{L}$ containing $\mathfrak{L}'$ is an ideal, and we conclude that $\mathfrak{L}$ has an ideal $\mathfrak{L}_{n-1}$ of dimension $n - 1$. The algebra $\mathfrak{L}_{n-1}$ is also solvable if $n - 1 > 0$; it has itself an ideal $\mathfrak{L}_{n-2}$ of dimension $n - 2$, and we may continue the process. Hence we have a chain $\mathfrak{L}_0 = \{0\} \subset \mathfrak{L}_1 \subset \mathfrak{L}_2 \subset \cdots \subset \mathfrak{L}_n = \mathfrak{L}$ of sub-algebras of $\mathfrak{L}$, where $\mathfrak{L}_i$ is of dimension i and contained as an ideal in $\mathfrak{L}_{i+1}$ (although $\mathfrak{L}_i$ is in general not an ideal in $\mathfrak{L}$). If for every i $(1 \leqq i \leqq n)$ we choose an element L_i belonging to $\mathfrak{L}_i$ but not to $\mathfrak{L}_{i-1}$, the n elements $L_1, L_2, \cdots, L_n$ constitute a base for $\mathfrak{L}$.

Suppose now that $\mathfrak{L}$ is the Lie algebra of a connected simply connected group G. Then every element in G can be written in one and only one way in the form $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_nL_n)$, where $t_1, t_2, \cdots, t_n$ are real numbers.[2] Let $\mathfrak{M}$ be any linear sub-space of $\mathfrak{L}$. We may choose the elements $L_1, L_2, \cdots, L_n$ in such a way as to satisfy the following condition: if there exists in $\mathfrak{L}_i$ an element of $\mathfrak{M}$ which does not belong to $\mathfrak{L}_{i-1}$, L_i shall be chosen in $\mathfrak{M}$. If $L_{i_1} = M_1, L_{i_2} = M_2, \cdots, L_{i_r} = M_r$ are the elements L_i which belong to $\mathfrak{M}$, the elements $M_1, M_2, \cdots, M_r$ obviously form a base for $\mathfrak{M}$.

Suppose now that $\mathfrak{M}$ is a sub-algebra. Let k, l be two indices such that $1 \leqq k < l \leqq r$; then $[M_k, M_l]$ is a linear combination of $M_1, M_2, \cdots, M_r$; on the other hand, it is also a linear combination of $L_{i_1}, L_{i_2}, \cdots, L_{i_{l-1}}$; hence it is a linear combination of $M_1, M_2, \cdots, M_{l-1}$. It follows that $M_1, M_2, \cdots, M_l$ constitute a base of a sub-algebra $\mathfrak{M}_l$ of $\mathfrak{M}$ of dimension l, and that $\mathfrak{M}_{l-1}$ is contained as an ideal in $\mathfrak{M}_l$: the base $M_1, M_2, \cdots, M_r$ has with respect to $\mathfrak{M}$ the same properties as the base $L_1, L_2, \cdots, L_n$ with respect to $\mathfrak{L}$.

Suppose that $\mathfrak{L}$ is the Lie algebra of a simply connected group G. Then the elements $(\exp u_1M_1)(\exp u_2M_2) \cdots (\exp u_rM_r)$ ($u_1, u_2, \cdots, u_r$ real) constitute a closed sub-set M of G, homeomorphic to the r-dimensional cartesian space. On the other hand, we know that there exists an abstract Lie group M' whose Lie algebra is $\mathfrak{M}$ and that every element in M' can be written in one and only

[4] Or *integrable* in the sense of S. Lie or E. Cartan.

one way in the form $(\exp u_1'M_1) \cdots (\exp u_r'M_r)$. It follows at once that M is a sub-group of G.

Therefore: *If G is a simply connected solvable Lie group, to every sub-algebra $\mathfrak{M}$ of the Lie algebra of G there corresponds a closed simply connected sub-group M of G whose Lie algebra is $\mathfrak{M}$.*

III. Study of a Particular Type of Solvable Group

Let G be a simply connected solvable Lie group which has the following properties:[5]

1) The Lie algebra $\mathfrak{L}$ of G, of dimension n, has an abelian ideal $\mathfrak{A}$ of dimension $n - 1$.

2) There exists an element σ of the center of G which does not belong to the connected sub-group $\mathfrak{C}$ of G whose Lie algebra is $\mathfrak{A}$.

We shall first prove that there exists an element $L \in \mathfrak{L}$ such that $\exp L = \sigma$. In fact the group G/A is an abelian connected group of dimension 1; hence there exists an element $\overline{L}$ of $\mathfrak{L}/\mathfrak{A}$ such that $\exp \overline{L} = \bar{\sigma}$, where $\bar{\sigma}$ is the coset modulo $\mathfrak{C}$ which contains σ. If L_1 is any element of $\mathfrak{L}$ whose coset modulo $\mathfrak{A}$ is $\overline{L}$, we have $\exp L_1 = \sigma\eta$, $\eta \in \mathfrak{C}$. If t is any real number, $\exp tL_1$ commutes with $\exp L_1$ and with σ, hence also with η. On the other hand, $\mathfrak{C}$ being abelian and connected, there exists one and only one element $A \in \mathfrak{A}$ such that $\eta = \exp A$. In the adjoint representation, the operation which corresponds to $\exp tL_1$ changes A into an element A_t such that $\exp A_t = \eta = \exp A$; hence $A_t = A$, which means that $[AL_1] = 0$; it follows that $\exp(L_1 - A) = \exp L_1 (\exp - A) = \sigma$; we may take $L = L_1 - A$.

Consider now the group $\mathfrak{G}$ composed of the elements $\exp tL$. Since $\exp L$ belongs to the center, the adjoint representation maps $\mathfrak{G}$ onto a group of linear operators in $\mathfrak{L}$ which is compact. The operators of this group permute among themselves the elements of $\mathfrak{A}$; since the group is compact, we have complete reducibility,[6] i.e. we can find a decomposition $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \cdots + \mathfrak{A}_h$ of $\mathfrak{A}$ into a direct sum of linear spaces $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_h$ each one of which gives an irreducible representation of $\mathfrak{G}$. Since $\mathfrak{G}$ is a one dimensional group (isomorphic to the additive group of real numbers), each $\mathfrak{A}_i$ is of dimension 1 or 2; suppose that $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_k$ are of dimension 2, $\mathfrak{A}_{k+1}, \cdots, \mathfrak{A}_h$ being of dimension 1.

The operation A_L which corresponds to L in the adjoint representation permutes among themselves the elements of $\mathfrak{A}_i$; let λ_i be one of its characteristic roots in $\mathfrak{A}_i$. Then e^{λ_i} is a characteristic root of the operation which represents $\exp L = \sigma$; hence $e^{\lambda_i} = 1$, $\lambda_i = 2\pi n_i\sqrt{-1}$, where n_i is an integer. If $n_i \neq 0$, λ_i is not real, and $-\lambda_i$ is also a characteristic root of A_L in $\mathfrak{A}_i$; hence $\dim \mathfrak{A}_i = 2$; if $n_i = 0$ on the contrary, we have $\dim \mathfrak{A}_i = 1$.

---

[5] The simplest example of such a group is the simply connected covering group of the group of motions in the plane.

[6] Cf. *Pontrjagin*, l. c. 2), Theorem 23, p. 110 and I, p. 109.

Moreover, if $i \leqq k$ we can find a base A_i, B_i of $\mathfrak{A}_i$ such that $\mathrm{A}_L(A_i + \sqrt{-1}\, B_i) = 2\pi n_i \sqrt{-1}\,(A_i + \sqrt{-1}\, B_i)$, i.e.

$$[A_i, L] = -2\pi n_i B_i$$
$$[B_i, L] = 2\pi n_i A_i \qquad (n_i \neq 0).$$

If $i > k$, we have

$$[A_i, L] = 0.$$

Hence the structure of $\mathfrak{L}$ is entirely determined when the integers n_i are known; the linear space $\mathfrak{A}_{k+1} + \cdots + \mathfrak{A}_h$ is the center of $\mathfrak{L}$, and $\mathfrak{A}_1 + \mathfrak{A}_2 + \cdots + \mathfrak{A}_k$ is the derived algebra of $\mathfrak{L}$.

It is easy to build from these data a system of finite equations for the group. We have n variables t, u_i, v_i $(1 \leqq i \leqq k)$, w_i $(k < i \leqq h)$ which constitute a system of coordinates on the group; σ_1, σ_2 being any two elements of the group, the law of composition is defined by the formulas

$$t(\sigma_1, \sigma_2) = t(\sigma_1) + t(\sigma_2) \qquad w_i(\sigma_1\sigma_2) = \exp\left(-2\pi n_i \sqrt{-1}\, t(\sigma_2)\right) \qquad (w_i(\sigma_1) + w_i(\sigma_2))$$

where we have set $w_i = u_i + \sqrt{-1}\, v_i$ if $1 \leqq i \leqq k$.[7] The coordinates of σ are $(1, 0, \cdots, 0)$.

We can also find the explicit formulas for the coordinates of $\exp X$, if $X = a_0 L + \sum_{i=1}^{k} (a_i A_i + b_i B_i) + \sum_{k+1}^{h} a_i A_i$: we find

$$t(\exp X) = a_0$$

if $1 \leqq i \leqq k$

$$w_i(\exp X) = \begin{cases} (a_i + \sqrt{-1}\, b_i)\dfrac{1 - \exp\left(-2\pi n_i \sqrt{-1}\, a_0\right)}{2\pi n_i \sqrt{-1}\, a_0} & \text{if } n_i a_0 \neq 0 \\ a_i + \sqrt{-1}\, b_i & \text{if } n_i a_0 = 0 \end{cases}$$

if $i > k$ $\qquad w_i(\exp X) = a_i.$

Hence, the elements X for which the equality $\exp X = \sigma$ holds are those of the form $L + \sum_1^k (a_i A_i + b_i B_i)$, and only those.

We shall need later the following consequence of these facts: let τ be any automorphism of the group G which leaves unchanged the element σ; if we denote by A_τ the linear operation[8] on $\mathfrak{L}$ which corresponds to τ, A_τ permutes among themselves the elements of the sub-algebra $\mathfrak{M}$ generated by L and the elements A_i, B_i for $1 \leqq i \leqq k$; if X is any element of $\mathfrak{M}$ such that $X \equiv L \pmod{\mathfrak{A}}$, we have again $\exp X = \sigma$.

---

[7] In order to simplify the writing of formulas, we are using complex valued functions w_i, although the group under consideration is of course a *real* Lie group.

[8] This operation is defined just as in the case of the adjoint-representation.

IV. The Central Theorem

We want to prove the following result:

Theorem 1. *Let G be a connected simply connected solvable group, with the Lie algebra $\mathfrak{L}$. Let D be a discrete sub-group of the center of G. Then D is a free abelian group of a certain rank r at most equal to the dimension n of G. It is possible to choose a base $L_1, L_2, \cdots, L_n$ of $\mathfrak{L}$ which has the following properties:*

1) *Every element in G can be written in one and only one way in the form* $(\exp t_1L_1)(\exp t_2L_2)\cdots(\exp t_nL_n)$.

2) *There exists a subset $\{i_1, i_2, \cdots, i_r\}$ of the set $\{1, 2, \cdots, n\}$ such that the elements* $\exp L_{i_1}, \exp L_{i_2}, \cdots, \exp L_{i_r}$ *constitute a base for D, and that* $[L_{i_\alpha}, L_{i_\beta}] = 0 (1 \leqq \alpha, \beta \leqq r)$.

We shall prove this theorem by induction on the dimension n of $\mathfrak{L}$. Let us assume that it has been proved already for all dimensions $< n$.

Since $\mathfrak{L}$ is solvable, it has certainly an abelian ideal of dimension > 0; let $\mathfrak{A}$ be an abelian ideal of maximal dimension; we set $\mathfrak{L}^* = \mathfrak{L}/\mathfrak{A}$. Let A be the connected abelian sub-group of G whose Lie algebra is $\mathfrak{A}$; A is an invariant sub-group, and G/A is a solvable group of dimension $< n$. The group DA/A is a sub-group of the center of G/A.

Consider now the closure $\overline{DA}$ of the group DA. The connected component of the unit element ϵ in this closure is a closed sub-group of G; according to a well known theorem of E. Cartan, it is a Lie group; let $\mathfrak{A}'$ be its Lie algebra. The group DA being abelian and invariant, $\mathfrak{A}'$ is an abelian ideal in $\mathfrak{L}$ containing $\mathfrak{A}$; in virtue of our choice of $\mathfrak{A}$, we have $\mathfrak{A}' = \mathfrak{A}$, which shows that the connected component of ϵ in $\overline{DA}$ reduces to A. On the other hand, there is a neighborhood of ϵ in $\overline{DA}$ which is compact; this neighborhood can meet only a finite number of distinct connected components of $\overline{DA}$, from which it follows at once that there exists a neighborhood U of ϵ in G such that $U \cap \overline{DA} \subset A$. Any element of DA which belongs to the set UA already belongs to A; hence the group DA/A is a discrete sub-group of G/A. It follows that we can apply our theorem to G/A and DA/A: we obtain a base $\{L_1^*, L_2^*, \cdots, L_m^*\}$ of $\mathfrak{L}/\mathfrak{A}$ with the following properties:

A) Every element of G/A can be written in one and only one way in the form $(\exp t_1L_1^*)(\exp t_2L_2^*)\cdots(\exp t_mL_m^*)$.

B) There exists a subset $\{i_1, i_2, \cdots, i_s\}$ of the set $\{1, 2, \cdots, m\}$ such that the elements $\exp L_{i_1}^*, \exp L_{i_2}^*, \cdots, \exp L_{i_s}^*$ constitute a base for the free discrete group DA/A, and that $[L_{i_\alpha}^*, L_{i_\beta}^*] = 0$ $(1 \leqq \alpha, \beta \leqq s)$.

For each α, we can find an element $\sigma_\alpha \in D$ such that its coset σ_α^* modulo A coincides with $\exp L_{i_\alpha}^*$.

We want now to choose in a certain way elements $L_1, L_2, \cdots, L_m$ of $\mathfrak{L}$ which will have to satisfy the following conditions: A) The coset of L_i modulo $\mathfrak{A}$ is L_i^* $(1 \leqq i \leqq m)$; B) we have $\exp L_{i_\alpha} = \sigma_{i_\alpha}$ $(1 \leqq \alpha \leqq s)$; C) we have $[L_{i_\alpha}, L_{i_\beta}] = 0$ $(1 \leqq \alpha, \beta \leqq s)$.

Suppose that we have already chosen $L_1, L_2, \cdots, L_h$ in such a way as to

satisfy the conditions of the type A), B), C) which concern these elements. We want to choose L_{h+1} (If $h = 0$, we assume simply that no choice has been made yet).

If $h + 1$ does not occur among the indices $i_1, i_2, \cdots, i_s$ we choose for L_{h+1} any element of $\mathfrak{L}$ whose coset modulo $\mathfrak{A}$ is L^*_{h+1}.

Suppose now that $h + 1 = i_{t+1}$. We begin by choosing an L'_{h+1} whose coset modulo $\mathfrak{A}$ is L^*_{h+1}, and we consider the Lie algebra $\mathfrak{H}$ spanned by $\mathfrak{A}$ and L'_{h+1}; let H be the corresponding connected subgroup of G. It is a closed simply connected sub-group of G. Moreover $\exp L'_{h+1}$ lies in the coset $\exp L^*_{h+1}$ modulo A; hence $\exp L'_{h+1} \equiv \sigma_{i_{t+1}}$ modulo A, which proves that $\sigma_{i_{t+1}} \in H$; since $\sigma_{i_{t+1}} \notin A$, H is a group of the kind which has been considered in §III. It follows first that $\mathfrak{H}$ contains an element L''_{h+1} such that $L''_{h+1} \equiv L'_{h+1}$ (modulo $\mathfrak{A}$) and that $\exp L''_{h+1} = \sigma_{i_{t+1}}$. Since we have $[L^*_{i_\alpha}, L^*_{h+1}] = 0$ $(1 \leqq \alpha \leqq s)$, we have $[L_{i_\gamma}, L''_{h+1}] \in \mathfrak{A}$ $(1 \leqq \gamma \leqq t)$. Since $[L_{i_\gamma}, L_{i_{\gamma'}}] = 0$ for $1 \leqq \gamma, \gamma' \leqq t$, the elements $(\exp u_1 L_{i_1}) \cdots (\exp u_{i_t} L_{i_t})$ constitute an abelian sub-group of G. Since $\mathfrak{A}$ is an ideal and $[L_{i_\gamma}, L''_{h+1}] \in \mathfrak{A}$, they are represented in the adjoint representation by linear operations which permute among themselves the elements of $\mathfrak{H}$ and $\mathfrak{A}$. The elements $\exp L_{i_\gamma}$ $(1 \leqq \gamma \leqq t)$ being represented by the unit operation, we obtain in this way a compact abelian group $\mathfrak{G}$ of linear transformations of $\mathfrak{H}$ into itself.

It follows from §III that $\mathfrak{H}$ can be represented in the form $\mathfrak{H}_1 \times \mathfrak{Z}$, where $\mathfrak{Z}$ is in the center of $\mathfrak{H}$ and $L''_{h+1} \in \mathfrak{H}_1$. Moreover the decomposition has, as we saw, the two further properties that: 1) if $X \in \mathfrak{H}_1$, $X \equiv L''_{h+1} (\mathrm{mod}\ \mathfrak{A})$, we have $\exp X = \exp L''_{h+1} = \sigma_{i_{t+1}}$ 2) any automorphism of H which leaves $\sigma_{i_{t+1}}$ invariant gives a linear operation in $\mathfrak{H}$ which permutes among themselves the elements of $\mathfrak{H}_1$.

This applies in particular to the inner automorphisms of G produced by the elements $(\exp u_1 L_{i_1})(\exp u_2 L_{i_2}) \cdots (\exp u_t L_{i_t})$. Hence the operations of the group $\mathfrak{G}$ permute among themselves the elements of $\mathfrak{H}_1$.

But $\mathfrak{H}_1 \cap \mathfrak{A}$ is a linear sub-space of $\mathfrak{H}_1$ whose dimension is exactly one less than the dimension of $\mathfrak{H}_1$. Its elements are also permuted among themselves by the operations of $\mathfrak{G}$. Since $\mathfrak{G}$ is compact, it is completely reducible, and we may conclude that there exists an element $L_{h+1} \in \mathfrak{H}_1$ with the following properties: 1) $L_{h+1} \equiv L''_{h+1}$ (mod $\mathfrak{A}$); 2) the elements of $\mathfrak{G}$ permute among themselves the multiples tL_{h+1} of L_{h+1}. Again because $\mathfrak{G}$ is compact, it follows that its operations leave L_{h+1} invariant; if we remember the definition of $\mathfrak{G}$, we see that $[L_{i_\gamma}, L_{h+1}] = 0$ $(1 \leqq \gamma \leqq t)$. Finally, the conditions $L_{h+1} \in \mathfrak{H}_1$, $L_{h+1} \equiv L''_{h+1}$ (mod $\mathfrak{A}$) imply $\exp L_{h+1} = \exp L''_{h+1} = \sigma_{i_{t+1}}$. Hence the element L_{h+1} satisfies all our requirements.

In this way, we shall be able to select m elements $L_1, L_2, \cdots, L_m$ satisfying the conditions A, B, C.

On the other hand, the group A is isomorphic to the additive group of R^{n-m}, and $D \cap A$ is a discrete sub-group of A, i.e. a lattice. It follows that we can find elements $L_{m+1}, \cdots, L_n$ in $\mathfrak{A}$ such that every element of A can be written

in one and only one way in the form $(\exp t_{m+1}L_{m+1}) \cdots (\exp t_nL_n)$ and that there exists a subset $\{i_{s+1}, \cdots, i_r\}$ of the set $\{m+1, \cdots, n\}$ such that the elements $\exp L_{i_{s+1}}, \cdots, \exp L_{i_r}$ constitute a base for $D \cap A$.

It remains only to prove that the sets $\{L_1, L_2, \cdots, L_n\}$ and $\{i_1, i_2, \cdots, i_r\}$ which have been selected have the properties 1), 2) of our theorem.

Let σ be any element of G and σ^* be its coset modulo A. Then we can write σ^* in the form $(\exp t_1L_1^*) \cdots (\exp t_mL_m^*)$. But this product represents a coset modulo A which contains the element $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_mL_m)$ since L_m belongs to the coset L_m^* modulo $\mathfrak{A}$. Hence we have

$$\sigma = (\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_mL_m)\zeta \qquad \zeta \in A \tag{1}$$

and ζ can be written in the form $(\exp t_{m+1}L_{m+1}) \cdots (\exp t_nL_n)$, which proves that σ is equal to an element of the form $(\exp t_1L_1)(\exp t_2L_2) \cdots (\exp t_nL_n)$. Moreover $t_1, t_2, \cdots, t_m$ are uniquely determined by σ^* (and *a fortiori* by σ); hence ζ is determined by σ and also $t_{m+1}, \cdots, t_n$.

If $\sigma \in D$, we have $\sigma^* \in DA/A$ and therefore σ^* can be written in the form $(\exp m_{i_1}L_{i_1}^*) \cdots (\exp m_{i_s}L_{i_s}^*)$, with integral $m_{i_1}, \cdots, m_{i_s}$. Since the element $(\exp m_{i_1}L_{i_1}) \cdots (\exp m_{i_s}L_{i_s}) = \sigma_{i_1}^{m_1}\sigma_{i_2}^{m_{i_2}} \cdots \sigma_{i_s}^{m_{i_s}}$ belongs to D, the element ζ of formula (1) belongs to D and can be written in the form $(\exp m_{i_{s+1}}L_{i_{s+1}}) \cdots (\exp m_{i_r}L_{i_r})$, which proves that the elements $\exp L_{i_1}, \exp L_{i_2}, \cdots, \exp L_{i_r}$ constitute a base for D.

We have already seen that $[L_{i_\alpha}, L_{i_\beta}] = 0$ for $1 \leqq \alpha, \beta \leqq s$. We shall see that, for $\alpha > s$, L_{i_α} necessarily belongs to the center of $\mathfrak{L}$. In fact, let us consider an arbitrary operation of the adjoint representation of G; it changes L_{i_α} in an element L'_{i_α} which still belongs to $\mathfrak{A}$; since $\sigma_\alpha = \exp L_{i_\alpha}$ belongs to the center we have $\exp L'_{i_\alpha} = \exp L_{i_\alpha}$. Since A is an abelian simply connected group, we conclude that $L'_{i_\alpha} = L_{i_\alpha}$, which proves that L_{i_α} belongs to the center of $\mathfrak{L}$. This completes the proof of our central theorem.

REMARK. We may assume that the set $\{i_1, i_2, \cdots, i_r\}$ is the set $\{1, 2, \cdots, r\}$. We again prove it by induction on n. Suppose that we have $\{i_1, i_2, \cdots, i_s\} = \{1, 2, \cdots, s\}$. We have seen that $L_{i_{s+1}}, \cdots, L_{i_r}$ belong to the center of $\mathfrak{L}$. It follows that any element $\exp t_{i_\alpha}L_{i_\alpha}$ $(\alpha > s)$ belongs to the center of G. Hence the property 1 stated in the theorem will still hold if we operate a permutation of $L_1, L_2, \cdots, L_n$ which brings L_{i_α} at α-th place $(\alpha = s+1, \cdots, r)$, and this proves our assertion.

THEOREM 9. *Let G^* be a connected solvable Lie group. There exists a compact abelian sub-group T of G^* and a sub-set E of G^*, homeomorphic to a cartesian space in a certain number of dimensions, such that every element $\sigma \in G^*$ can be written in one and only one way in the form $\sigma = \tau\eta$, $\tau \in T$, $\eta \in E$. The mapping $(\tau, \eta) \to \tau\eta$ is a homeomorphism of $T \times E$ with G^*.*

In fact we have $G^* = G/D$ where G is a simply connected solvable Lie group and D a discrete sub-group of the center of G. We can find a base $\{L_1, L_2, \cdots, L_n\}$ of the Lie algebra of G such that every element in G can be written in one and only one way in the form $(\exp t_1L_1) \cdots (\exp t_nL_n)$, and that

the elements $\exp L_1$, $\exp L_2$, $\cdots$, $\exp L_r$ constitute a base for D. Moreover the group A of the elements $(\exp t_1L_1) \cdots (\exp t_rL_r)$ is abelian.

We have $D \subset A$ and A/D is a compact abelian sub-group T of G^*.

The elements $(\exp t_1L_1) \cdots (\exp t_nL_n)$ and $(\exp t_1'L_1) \cdots (\exp t_n'L_n)$ are congruent modulo D if and only if $t_1' \equiv t_1 \pmod 1$, $\cdots$, $t_r' \equiv t_r \pmod 1$, $t_{r+1}' = t_{r+1}$, $\cdots$, $t_n' = t_n$. Let φ be the natural mapping of G onto G^*. The set $\bar{E}$ of the elements of the form $(\exp t_{r+1}L_{r+1}) \cdots (\exp t_nL_n)$ is mapped by φ in one-to-one continuous way onto a subset E of G^*. We claim that this mapping is a homeomorphism. In fact, let $\bar{U}$ be the set of the elements $(\exp t_1L_1) \cdots (\exp t_nL_n)$ with $|t_1| < \frac{1}{2}$, $\cdots$, $|t_r| < \frac{1}{2}$; $\bar{U}$ is a neighborhood of $\bar{E}$ in G and is mapped in a one-to-one continuous way by φ; but we know that φ maps any open subset of G onto an open subset of G^*; hence φ gives a homeomorphism of U with a subset of G^*, which proves our assertion.

It follows that E is homeomorphic to the $(n - r)$-dimensional cartesian space. Moreover every element of G^* can obviously in one and only one way be written in the form $\tau\eta$, $\tau \in T$, $\eta \in E$.

Since the mapping $(\tau, \eta) \to \tau\eta$ clearly maps open sets of $T \times E$ onto open sets of G^*, it is a homeomorphism, and theorem 2 is completely proved.

Since a compact abelian group is a torus-space (product of circumferences) we have the

THEOREM 2^a. *Any connected solvable Lie group is homeomorphic to the product of a torus space and of a cartesian space.*

PRINCETON UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ALTERNATIVE RINGS AND RELATED QUESTIONS I: EXISTENCE OF THE RADICAL

BY MAX ZORN

(Received July 29, 1940)

In a previous publication (1) I have presented a theory of alternative rings which led up to the characterization of semisimple systems. The present paper is concerned with general hypercomplex[1] alternative rings; its main object is to establish the existence of the radical. In other words it is shown that the properly nilpotent elements form a (two-sided) ideal.

The first section contains the proof of several alternative identities,[2] which are then applied (sec. 2) to the theory of the inverse element. Though a certain amount of duplication[3] could not be avoided, the new arrangement has been given in detail because it is phrased almost entirely in terms of "associators" $(a, b, c) = (ab)c - a(bc)$. A new proof of the O-lemma[4] has likewise been included; the lemma is now obtained by inspection of an identity, and the assumption about characteristics is removed. This is of interest in connection with the theory of Peirce decompositions, which was derived in (1) under the (now unnecessary) restriction that the characteristics be different from 2.

Thus far the theory is entirely self-contained; this is likewise achieved in the third section by formulating two theorems from (1) as assumptions. These two propositions are indeed so simple that they may very well serve as axioms. On this basis it is possible to treat the rings which in the associative theory have been called "vollstaendig primaer";[5] the existence of the radical is verified in this important special case.

In the last two sections use is made of the theory of idempotents and Peirce decompositions as developed in (1); the necessary material is very briefly reviewed in sec. 4. A certain ideal N is constructed in the fifth section, which depends formally on an arbitrary but fixed maximal Peirce decomposition. The concluding section contains the proof that this ideal consists exactly of all properly nilpotent elements and may therefore be regarded as the radical of the ring.

NOTATIONS. With one exception (the delta in the Kronecker symbol δ_{ij}) we use only italics; i, j, k, l, m, n for indices and the like, $a, b, c, d, e, x, y_{ik}, z_{ii}$ for elements, capitals $S, S_{ik}, N_{ik}, N, A, B$ for sets.

---

[1] See sec. 4 and also (1) p. 137.

[2] I am indebted to P. Jordan for many remarks concerning such and more general identities.

[3] See (1) p. 142.

[4] See (1) pp. 129, 130.

[5] See (2) p. 256 and (3) p. 85.

The summation sign is used in such a way that the range of the index is described in a parenthesis which precedes the letter $\sum$.

Theorems and definitions are numbered decimally within the section and the number of the section is prefixed; a theorem is indicated by a parenthesis—(3.7.1)—and a definition by a bracket—[1.1]—. A single numeral in a parenthesis, like (2), refers to the bibliography.

1. We shall constantly have to use the abbreviation:

$$[1.1] \qquad (a, b, c) = (ab)c - a(bc).$$

These "associators" satisfy in every distributive ring the identity

$$(1.2) \qquad (ab, c, d) - (a, bc, d) + (a, b, cd) = a(b, c, d) + (a, b, c)d$$

which becomes trivial when the associators are worked out according to [1.1].

Let us permute the arguments cyclically such that b is the first letter, and add the resulting equation to (1.2); we get

$$(1.3) \quad (ab, c, d) + (b, c, da) + [(bc, d, a) - (a, bc, d)] + [(a, b, cd) - (b, cd, a)] = a(b, c, d) + (a, b, c)d + b(c, d, a) + (b, c, d)a.$$

From now on we ask that multiplication in the distributive ring satisfy the alternative law, which can be expressed in different ways:

[1.4] A ring is called *alternative* and said to satisfy the alternative law of multiplication if the associator (a, b, c) is an alternating function of its arguments. The associator must therefore be zero if two of the arguments are equal; it takes on the factor -1 if two of the arguments are transposed and it remains unchanged if the arguments undergo a cyclic permutation:

$$(1.4.1) \quad (a, a, b) = (a, b, a) = (b, a, a) = 0,$$

$$(a, b, c) + (b, a, c) = (a, b, c) + (a, c, b) = (a, b, c) + (c, b, a) = 0,$$

$$(a, b, c) = (b, c, a) = (c, a, b).$$

The left hand member of the equation (1.3) would consequently reduce to $(ab, c, d) + (b, c, da)$; if we put $b = c$, most of the terms will drop out and the equation $(ab, b, d) = b(b, d, a)$ remains. Exchanging a and b and applying the alternative law we get the identity

$$(1.6) \qquad (a, ba, c) = -a(b, a, c).^{6}$$

We add this equation to

$$(ab, a, c) - (a, ba, c) + (a, b, ac) = a(b, a, c) + (a, b, a)c$$

---

[6] (1) p. 142.

such as to get

$$(ab, a, c) + (a, b, ac) = 0. \tag{1.7}$$

On this identity we shall base, in the following section, the theory of the inverse element.

From (1.7) we pass to the next identity

$$(ab, a, c) + (ba, a, c) - (b, a^2, c) = 0 \tag{1.8}$$

by way of adding

$$(ba, a, c) - (b, a^2, c) + (b, a, ac) = b(a, a, c) + (b, a, a)c = 0.$$

2. The theory of Peirce decompositions, which was given in (1), rests on the "O-lemma":

(2.1) *If $ab = ba = 0$ then for every c the associator (a^2, b, c) vanishes.*

Without any restriction this lemma follows by substitution from the identity (1.8); thus we are entitled to apply all results of (1) about idempotent elements without reservation. This will be of importance in the concluding sections.

For the remainder of this section we deal with the theory of the inverse and assume consequently the existence of a unit element.

Once more we apply the identity (1.8), this time for $ab = ba = 1$:

(2.2) *If $ab = ba = 1$, then all associators of the forms (a^2, b, c) and (a, b^2, c) are zero.*

In order to extend this to associators of the type (a, b, c) we replace b by b^2 in the formula (1.7):

$$(ab^2, a, c) + (a, b^2, ac) = 0. \tag{2.3}$$

Since in accordance with the alternative law the equation $ab = 1$ implies that $ab^2 = a(bb) = (ab)b = 1b = b$, we get from (2.2) and (2.3) the theorem:

(2.4) *If the numbers a and b satisfy the equations $ab = ba = 1$, then every associator of the form (a, b, c) vanishes.*[7]

The element b (if it exists) has all properties of an inverse element for a; it is unique and may be used for the solution of linear equations:

(2.5) *If $ab = ba = 1$, then the equations $ax = c$ and $ya = c$ have the solutions $x = bc$ and $y = cb$; these solutions are unique.*

Indeed, the preceding theorem permits us to proceed as is customary for associative systems. The existence of the solutions follows from formulas like $a(bc) = (ab)c - (a, b, c) = 1c - 0 = c$, and the uniqueness is exhibited by similar equations like $cb = (ya)b = y(ab) + (y, a, b) = y1 + 0 = y$.

In particular, by considering the cases $c = 1$ and $c = 0$, we obtain the corollary:[8]

(2.6) *If $ab = ba = 1$, then b is the unique solution of either one of the two equations $ax = 1$ or $ya = 1$; the element a is not a divisor of zero.*

---

[7] Compare (1) p. 142, lemma 2.

[8] For additional information about the inverse in alternative fields see (4).

This justifies the definition:

[2.7] If there exists an element b satisfying $ab = ba = 1$, then it is called the *inverse* of a and denoted by a^{-1}.

This definition and the foregoing results permit us to state:

(2.8) $$aa^{-1} = a^{-1}a = 1; \qquad (a^{-1})^{-1} = a; \qquad (a, a^{-1}, c) = 0.$$

3. In order to postpone as far as possible any reference to outside material we introduce the following two assumptions.

(3.1) *Any two elements a, b of an alternative ring generate, by multiplications and additions, an associative subring.*

That this assumption is satisfied in every alternative ring has been shown by Artin;[9] one particular consequence is that the powers a^n with positive exponents are commutative and associative with each other.

(3.2) *An element a is either nilpotent, or else there exists a right multiple $e = ab$ which is idempotent, such that $e^2 = e \neq 0$.*[10]

[3.3] DEFINITION. An alternative ring will be called *primitive* if (a) it contains a unit element 1 which is different from 0; (b) the unit element is the only idempotent of the ring; (c) it satisfies the two above assumptions (3.1) and (3.2).

In defining next the properly nilpotent elements we deviate from the classical definition[11] by including 0. This is done because otherwise we would always have to say "properly nilpotent or zero."

[3.4] An element a is called *properly nilpotent* if all multiples ax are nilpotent. By virtue of (3.1) we see that $(ax)^m = 0$ implies $(xa)^{m+1} = x(ax)^m a = 0$, such that also the left multiples of a are nilpotent; we shall have to use this remark later.

(3.5) *If a multiple ax has an inverse then the element a is not nilpotent.*[12]
Let $a^{m-1} \neq 0 = a^m$; then $0 \neq a^{m-1} = a^{m-1}((ax)(ax)^{-1}) = (a^{m-1}(ax))(ax)^{-1} = (a^m x)(ax)^{-1}$ shows that $a^m = 0$ is impossible.

The theorems (3.6)–(3.9) describe some important properties of primitive rings.

(3.6) *If an element a of a primitive ring is not nilpotent then it has an inverse.*
Indeed there must be a multiple ab which is idempotent, by virtue of (3.2). This idempotent must be equal to 1 since our ring is primitive; the theorem will be proved if we show that $ba = 1$.

First of all we see that ba is not equal to zero; for in that case $1 = 1^2 = (ab)(ab) = a(ba)b = 0$ would ensue.[13] Secondly ba is idempotent because of $(ba)(ba) = b(ab)a = (b1)a = ba$; consequently we have $ba = 1$ and b is the inverse of a.

---

[9] See (1) p. 127.

[10] This assumption appears in (1) p. 138 as a theorem, valid in hypercomplex rings and incidentally based on the first and a part of the second chain condition.

[11] See (3) p. 46.

[12] Compare (1) p. 142.

[13] This part was omitted in (1) p. 142.

(3.7) *Every nilpotent element of a primitive ring is properly nilpotent.*

(3.7.1) COROLLARY. *Every element of a primitive ring is either properly nilpotent or it has an inverse.*

In other words if a is nilpotent then every multiple is nilpotent. That is an immediate consequence of (3.5) and (3.6). For if—to conclude indirectly—one multiple ax were not nilpotent it would possess an inverse and (3.5) imply that a is not nilpotent. Using this and the remark about [3.4] we find:

(3.7.2) *If an element a of a primitive ring is properly nilpotent, every multiple ax and xa is properly nilpotent.*

Thus one of the two essential properties of two-sided ideals is verified for the set of all properly nilpotent elements. It remains to show that this set is a modulus (i.e. closed under addition and subtraction). This is accomplished if we prove the theorem:

(3.8) *If a and b are two properly nilpotent elements of a primitive ring, then the the difference $a - b$ is likewise properly nilpotent.*

An equivalent proposition would be:

(3.8.1) *If the element b of a primitive ring is properly nilpotent, and the difference $a - b$ is not, then the element a is not (properly) nilpotent.*

For in that case the difference $a - b$ has in inverse; let $(a - b)x = ax - bx = 1$, so that $ax = 1 + bx$. Since bx is nilpotent, with say $(bx)^m = 0$, the element $1 + bx$ has the inverse $(i = 0 \cdots m) \sum (-bx)^i$. Thus the element a is seen to have a multiple with an inverse; and that precludes, according to the theorem (3.5), the possibility that a be nilpotent.

If we observe finally that the set of all properly nilpotent elements is not empty we may enunciate the theorem:

(3.9) *The properly nilpotent elements of a primitive ring form a two-sided ideal.*

We call this ideal the radical of the primitive ring and define more generally:

[3.10] If the properly nilpotent elements of an alternative ring form an ideal, then this ideal is called the *radical* of the ring in question.

We note without proof:

(3.11) *The residue class ring modulo the radical (if it exists) is semisimple;*

(3.12) *The residue class ring modulo the radical of a primitive ring is an alternative field;*

and

(3.12.1) *If a primitive ring does not contain properly nilpotent elements different from zero, or in other words if the zero ideal is the radical, then the ring is an alternative field.*

We intend to establish the existence of a radical for all hypercomplex rings; for that reason we state as an immediate consequence of (3.2) and [3.10] the fact:

(3.13) *If a ring does not contain any idempotent elements, it possesses a radical, which coincides with the ring itself.*

4. In order to complete our proof we shall have to use some of the definitions and theorems which were developed in (1).

[4.1.1] The set of all right multiples ax of an element a (*I-modulus*) is denoted by aS;

[4.1.2] The solutions of $ax = 0$ form a "*II-modulus*" denoted by $(a\backslash 0)$;

[4.1.3] The modulus generated by the products ab where a and b are respectively taken from the moduli A and B is denoted by AB.

[4.2] An alternative ring is called *hypercomplex* if it satisfies the following three chain conditions:

(4.2.1) every (necessarily descending) chain a^iS is finite;

(4.2.2) every ascending chain of II-moduli $(a^i\backslash 0)$ is finite;

(4.2.3) every descending chain of II-moduli $(a_i\backslash 0)$ is finite.

For such rings the following theorems were derived:[14]

(4.3) *There exists (at least) one finite set $e_1 \cdots e_n$ of idempotents with $e_ie_k = 0$ whenever i and k are different, such that the ring S is the direct sum*

$$S = (i, k = 0 \cdots n) \sum S_{ik}$$

of moduli S_{ik}, the modulus S_{ik} consisting exactly of the elements x_{ik} which satisfy

$$e_jx_{ik} = \delta_{ji}x_{ik}, \qquad x_{ik}e_j = \delta_{kj}x_{ik};$$

the "component" of x in the modulus S_{ik} is denoted by x_{ik} and for $i \neq 0 \neq k$ is given by e_ixe_k; this "Peirce-decomposition" has the following properties (4.3.1) *and* (4.3.2).

(4.3.1) *The moduli S_{0i} and S_{i0}, for $i = 0 \cdots n$, consist of properly nilpotent elements; the modulus S_{00} is a subring.*

(4.3.2) *The moduli S_{ii}, for $i = 1 \cdots n$, are hypercomplex and primitive subrings; their respective unit elements are the idempotents e_i.*

The last two properties express the fact that the decomposition is a maximal one. Without loss of generality[15] we assume that there are actually idempotent elements, and that consequently the number n is not 0. Moreover it can be shown by a refinement of the construction in (1) that an additional condition can be satisfied, to wit

(4.4) *If e is an idempotent element, then the Peirce decomposition may be chosen in such a way that e is a sum of elements e_i, or*

$$e = (i = 1 \cdots m) \sum e_i.$$

The next theorem deals with the products of the moduli S_{ik}. Like the remaining statements from this section it is valid for general Peirce decompositions.

(4.5.1) $$S_{ik}S_{kl} \subset S_{il},$$

(4.5.2) $$S_{ik}S_{ik} \subset S_{ki},$$

[14] Under the restriction that $a + a = 0$ imply $a = 0$; we have removed this restriction in the second section.

[15] See (3.13) and (3.2)

(4.5.3) All other products $S_{ik}S_{lm}$ are zero.

The second of these statements is amplified by the formula:

$$(4.6.1) \qquad x_{ik}y_{ik} + y_{ik}x_{ik} = 0 \quad \textit{if} \quad i \neq k;$$

this fact is going to be used in the weaker form:

(4.6.2) *If N_{ik} is a modulus contained in S_{ik}, and if $i \neq k$, then we can say that $N_{ik}S_{ik} = S_{ik}N_{ik}$.*

Concerning products of three factors only a part of the "Dreierregeln"[16] is necessary for our purposes, which we report in the next theorem:

(4.7.1)[17] *With the exception of products like $(x_{ii}y_{ii})z_{ii}$ every product of type $(x_{ik}y_{kl})z_{li}$ is associative, that is, equal to $x_{ik}(y_{kl}z_{li})$.*

(4.7.2)[18] *A product of type $(x_{ik}y_{ik})z_{ik}$, where the indices i and k are different retains its value if the factors undergo a cyclic permutation.*

We name finally the rules[19] about powers of the form $(a_{ii}x)^m$ and $(a_{ik}x)^m$:

$$(4.8.1) \qquad (a_{ii}x)^m = (a_{ii}x_{ii})^{m-1}(j = 0 \cdots n) \sum a_{ii}x_{ij};$$

and for $i \neq k$

$$(4.8.2) \quad (a_{ik}x)^m = (a_{ik}x_{ki})^{m-1}(j = 0 \cdots n) \sum a_{ik}x_{kj} + (a_{ik}x_{ik})(a_{ik}x_{ki})^{m-1}.$$

If we compare, in the latter formula, the components in S_{ii} of the two members, we see that

(4.8.3) $e_i(a_{ik}x)^m e_i = (a_{ik}x_{ki})^m$, *and this holds likewise if the two indices are equal.*

The formulas (4.8) will only be used in the last section.

5. With the aid of theorems developed in 4 and reported about in 5 we shall now, for a fixed but arbitrary Peirce decomposition with the properties enumerated in section 4, construct an ideal which later on will turn out to be the

---

[16] See (1) p. 134.

[17] See (1) p. 136, last paragraph of sec. 3.

In order to obtain the theorem (4.7.1) from (1) it is necessary to distinguish various cases; the following direct proof seems more satisfactory. We see that if two of the ordered index pairs ij, jk, ki are equal, then the three indices are equal; for due to the cyclic symmetry of the situation we may assume in such a case that the first two pairs are identical, and that would mean $i = j$ and $j = k$. In those cases which are not classed as exceptions by the wording of the theorem we have to deal with three different pairs of indices, and we may consider the elements in question as components of a single element, namely, the sum $x_{ij} + y_{jk} + z_{ki} = a$. But the components of one element a with respect to a fixed Peirce decomposition are contained in the subring generated by a and the idempotents e_l.

A single element generates an A-ring (see (1) p. 127); and so does, by virtue of the O-lemma, and the A-lemma, any set of orthogonal elements. According to the generalized theorem of Artin two A-rings generate an associative ring, and thus the components of a single number are seen to be associative with each other. In particular the three numbers x_{ij}, y_{jk}, and z_{ki} are associative, q.e.d.

[18] See in (1) on p. 135 the more general formula (23).

[19] See (1) p. 136.

radical of the ring S. This ideal N will consist exactly of the quantities whose components are "singular" according to the following definition.

[5.1] An element a_{ik} (of S_{ik}) is called *singular* if all elements of $a_{ik}S_{ki}$ are nilpotent, otherwise regular.[20]

The following statements are easily deduced from the theorems which are indicated thereafter.

(5.2.1) *If one of the indices i or k is zero then every element a_{ik} is singular.* See (4.3.1).

(5.2.2) *For equal indices the elements a_{ii} constitute the radical of the ring S_{ii}.* See (4.3.2) and (3.13).

(5.2.3) *If a_{ik} is singular, all elements of $S_{ki}a_{ik}$ are likewise nilpotent.* See the remark about [3.4].

We define:

[5.3] The set of all singular elements in S_{ik} shall be denoted by N_{ik}.

We see that the sets N_{i0} and N_{0i} are identical with S_{i0} and S_{0i} respectively and are therefore moduli; likewise—according to (5.2.2)—the sets N_{ii} are the radicals of S_{ii} and are therefore moduli. We maintain

(5.4) *The sets N_{ik} are moduli in any case.*

Suppose that both elements a_{ik} and b_{ik} are singular, and consider the difference $a_{ik} - b_{ik}$. Every multiple $(a_{ik} - b_{ik})x_{ki}$ is the difference of two numbers, namely, $a_{ik}x_{ki}$ and $b_{ik}x_{ki}$, from the radical of S_{ii}; such differences are necessarily in the radical N_{ii}.

The definition of the sets N_{ik} implies that $N_{ik}S_{ki} \subset N_{ii}$, and from (5.2.3) we infer also that $S_{ik}N_{ki} \subset N_{ii}$. Again we make a statement which is slightly more general:

(5.5) *The products $N_{ij}S_{jk}$, $S_{ij}N_{jk}$, $N_{ki}S_{ki}$ and $S_{ki}N_{ki}$ are contained in the modulus N_{ik}.*

For i equal to k nothing new is asserted; for the other cases it is necessary and sufficient to verify the following inclusion relations:

$$(5.5\,1) \qquad (N_{ij}S_{jk})S_{ki} \subset N_{ii},$$

$$(5.5.2) \qquad (S_{ij}N_{jk})S_{ki} \subset N_{ii},$$

$$(5.5.3) \qquad (N_{ki}S_{ki})S_{ki} \subset N_{ii},$$

$$(5.5.4) \qquad (S_{ki}N_{ki})S_{ki} \subset N_{ii}.$$

Evidently the last of these formulas is equivalent to the third, on account of (4.6.2). According to (4.7.2) the third relation is equivalent to $(S_{ki}S_{ki})N_{ki} \subset N_{ii}$, *and this would follow* from the second formula because we know that $S_{ik}S_{ik} \subset S_{ki}$. The second inclusion in turn would be a consequence of the first. For a product of this type is associative,[21] therefore instead of (5.5.2) the rela-

[20] Compare (2) p. 257.

[21] See (4.7.1).

tion of $S_{ij}(N_{jk}S_{ki}) \subset N_{ii}$ may be discussed; and if indeed $N_{jk}S_{ki}$ is contained in N_{ji}, as is implied by the first formula in this group, then (5.5.2) must be true.

The first formula, on which the others are thus seen to depend, follows from $S_{jk}S_{ki} \subset S_{ji}$ if we remember that according to theorem (4.7.1) the product $(N_{ij}S_{jk})S_{ki}$ is associative.

We define now the set N which we claim to be the radical of the ring.

[5.6] The set of all elements whose components are singular with respect to the Peirce decomposition under discussion is denoted by N; in other terms $N = (i, k = 0 \cdots n) \sum N_{ik}$.

As announced in the beginning of this section, we have the theorem:

(5.7) *The set N is an ideal.*

First of all we see that by virtue of (5.4) the set N is a modulus; it remains to show that the products NS and SN are contained in N. In accordance with the rules (4.5) we have

$$NS = (\sum N_{ik})(\sum S_{lm}) = (i, j, k = 0 \cdots n) \sum N_{ij}S_{jk} + (i \neq k) \sum N_{ik}S_{ik};$$

the relations (5.5) have been derived in order to make evident that the last member is contained in N; the other relation $SN \subset N$ is verified in the same fashion.

We note that the ideal N does not contain any one of the elements e_i (these are regular) and is therefore different from the ring S. This would allow us to state another special instance of the final theorem, namely:

(5.8) *If a hypercomplex ring is simple and does not consist exclusively of nilpotent elements, then the only properly nilpotent element is zero. A fortiori, the 0-ideal is the radical of the ring.*

This partial result is of value in the theory of finite alternative division algebras.[22]

6. By means of the following lemma we show that every properly nilpotent element is contained in the ideal N.

(6.1) *If one component, say x_{ki}, of an element x is not singular, then we can find a left multiple $ax = a_{ik}x$ which is not nilpotent.*

Suppose that there is an element a_{ik} such that $a_{ik}x_{ki}$ is equal to e_i; the existence of such a quantity will be shown presently. We say that $a_{ik}x$ cannot be nilpotent, since the formula (4.8.3) makes it evident that the component in S_{ii} of this product and any of its powers is equal to e_i. In order to prove that the element a_{ik} exists we note first that none of the indices is zero, since x_{ki} was regular. If the indices are equal, then x_{ii} has an inverse in the ring S_{ii} which solves the equation. If finally i and k are different, we choose first an element y_{ik} such that $y_{ik}x_{ki}$ is not nilpotent; the latter product has again an inverse z_{ii} in the ring S_{ii}, so that $z_{ii}(y_{ik}x_{ki}) = e_i$. Since the last product is associative, by virtue of (4.7.1), we see that $a_{ik} = z_{ii}y_{ik}$ is the element we wanted.

---

[22] See (5) p. 402.

In the other direction we prove:

(6.2) *Every element of a modulus N_{ik} is properly nilpotent.*

Indeed, according to the formulas (4.8.1) and (4.8.2) the power $(a_{ik}x)^m$ is a sum of at most two terms which contain the factor $(a_{ik}x_{ki})^{m-1}$; the latter quantity vanishes for sufficiently high exponents, and we see that every multiple of a_{ik} is nilpotent. This implies in particular that the elements of the ideal N are at least sums of nilpotent elements, and we state less if we say:

(6.2.1) *The ideal N is contained in every ideal which contains all properly nilpotent elements.*

Combining this with theorem (6.1), which guarantees that every properly nilpotent element is necessarily a member of N, we obtain:

(6.2.2) *The ideal N is exactly the ideal which is generated by the properly nilpotent elements of the ring.*

This remark makes it clear that the ideal does not depend on the particular Peirce decomposition which has been used in the construction. We have mentioned before that the idempotents e_i are not elements of the ideal, since their components in S_{ii} are not singular; the same holds naturally for any sum of the form $(i = 1 \cdots m) \sum e_i$, provided that it is not zero. By virtue of (4.4) we may choose a Peirce decomposition for which a preassigned idempotent e turns out to be a sum of this type, and that implies:

(6.3) *The ideal N has no idempotent elements.*

If an ideal does not contain any idempotents then all its elements are nilpotent; for an element which is not nilpotent has an idempotent multiple; and an ideal contains the multiples of its elements. It is obvious, for the same reason, that all multiples of any element in the ideal are nilpotent. Thus we have finally arrived at the conclusion:

(6.4) *The ideal N, like any ideal of a hypercomplex ring without idempotents, consists entirely of properly nilpotent elements.*

If we combine the theorems (3.13), (6.1), (6.4) and the definitions [3.4] and [3.10] we may finally state the

(6.5) THEOREM. *Every hypercomplex alternative ring has a radical.*

Conclusion. If one is willing to restrict the class of hypercomplex rings the theory becomes much simpler. Mr. Artin has proved the existence of the radical for finite algebras of characteristic zero by extending the theory of traces.[23] By the use of an entirely different method (developed[24] first for Lie rings) I have been able to establish the nilpotency of the radical for alternative systems which satisfy the ascending chain condition for subsystems; this takes care of arbitrary finite algebras and likewise of rings which have a finite number of elements. Whether this fact is also a consequence of our three chain conditions I do not know.

---

[23] Not published; see (5) p. 402.

[24] See (6).

On the other hand it seems very probable that the three chain conditions are not the weakest assumptions which furnish a radical. Several intermediate results can be established under conditions which are compatible with infinite Peirce decompositions; and it seems that there is a less restrictive theory where infinite matrices with a finite number of nonvanishing rows play a role.

UNIVERSITY OF CALIFORNIA AT
LOS ANGELES

BIBLIOGRAPHY

1. M. Zorn, *Theorie der alternativen Ringe*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 8 (1930), p. 123.
2. E. Artin, *Zur Theorie der hyperkomplexen Zahlen*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 5 (1927), p. 251.
3. L. E. Dickson, *Algebras and their arithmetics*, Chicago 1923.
4. R. Moufang, *Zur Struktur von Alternativkörpern*, Mathematische Annalen, vol. 110 (1934), p. 416.
5. M. Zorn, *Alternativkörper und quadratische Systeme*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 9 (1933), p. 395.
6. M. Zorn, *On a theorem of Engel*, Bulletin of the American Mathematical Society, vol. 43 (1937), p. 401.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

SEPARATION SPACES[0]

BY A. D. WALLACE

(Received July 17, 1940)

1. Introduction

The purpose of this paper is to give a system of axioms suitable for the construction of a theory of connectivity in the large and for the study of connectivity invariants. The postulate system, which is of an essentially new type, is based on the notion of separation as the undefined concept. In sections 2 and 3 we correlate our system with the usual topological concepts, in section 4 proofs are given for the "expected" theorems, and sections 5 and 6 are devoted to some general examples. Section 7 is devoted to transformations and a generalization of the Eilenberg-Whyburn theorem on the factorization of continuous transformations.

2. Axioms and Definitions

We envisage an abstract set of elements S together with a binary relation $X \mid Y$ between subsets X and Y of S. By a *weak separation-space* (or an s_w-space) we mean a system $(S, \mid)$ in which the following axioms hold:

AXIOM S.I. *The null set is separated from every set*: $(X \subset S) \rightarrow (0 \mid X)$.[1]

AXIOM S.II. *Separation is symmetric*: $(X \mid Y) \rightarrow (Y \mid X)$.

AXIOM S.III. *Separation is disjunctive*: $(X \mid Y) \rightarrow (X \cdot Y = 0)$.

AXIOM S.IV. *Separation is hereditary*: $(X \mid Y) \cdot (X_1 \subset X) \rightarrow (X_1 \mid Y)$.

For convenience we adopt the conventions:

(i) The symbol "$X \mid Y$" is to be read "*X is separated from Y*" or "*X and Y are separated.*"

(ii) In a topological space two sets X and Y will be said to be "*topologically separated*" if they satisfy the condition of Lennes-Hausdorff, $\bar{X} \cdot Y + X \cdot \bar{Y} = 0$.

Before examining particular instances of s_w-spaces it will be necessary to have the

(2.1) DEFINITION. *If X is a subset of an s_w-space then kX is the set of all points no one of which is separated from X: $(p \in kX) \equiv (p \mid X)'$. By $H(X)$ is meant the class of all sets Y such that $Y = kY \supset X$, and hX is the set of points common to all the sets in the class $H(X)$.*

The topology arising from the adoption of kX (respectively hX) as the *closure* of the set X will be called the *k-topology* (respectively *h-topology*) of the s_w-space.

---

[0] Presented to the Am. Math. Soc., April 8, 1939 and April 26, 1940.

[1] We adopt the logical notation of Kuratowski [1]. Thus if A and B are propositions then $A \equiv B$, $A \rightarrow B$, $A + B$, $A \cdot B$, A' are respectively A is equivalent to B, A implies B, A or B, A and B, and A is false. The symbol $(X \mid Y)'$ will be constantly used.

(2.2) EXAMPLE. Let S be any set of elements and let "$X \mid Y$" mean that X and Y are disjoint. Here the system $(S, \mid)$ is an s_w-space in which $X = kX = hX$ so that the k-topology and the h-topology are each discrete.

(2.3) EXAMPLE. Let S be a topological space in which *closure* satisfies the axioms *K.I–K.III* of (3.3), i.e., S is a T_1-space. Define "$X \mid Y$" to mean that X and Y are topologically separated, $\bar{X} \cdot Y + X \cdot \bar{Y} = 0$. It is clear that the system $(S, \mid)$ is a weak separation-space with $\bar{X} = kX = hX$.

(2.4) EXAMPLE. In this example S is Euclidean n-space and "$X \mid Y$" is to mean the existence of an $(n-1)$-dimensional subspace L such that X and Y lie in different open half-spaces determined by L. Here we have $\bar{X} \subset hX = kX$, and this latter set is the closed convex hull of X. In this connection see §4.

(2.5) THEOREM. *In an s_w-space we have*

K.I$_w$ If $X \subset Y$ then $kX \subset kY$.

K.II$_w$ $X \subset kX$, $k0 = 0$.

H.I$_w$ If $X \subset Y$ then $hX \subset hY$.

H.II$_w$ $X \subset hX$, $h0 = 0$.

H.III$_w$ $hhX = hX$.

HK $X \subset kX \subset hX = hkX = khX$.

PROOF: (K.I$_w$). If $x \in kX$ then $(x \mid X)'$ and hence by S.IV it follows that $(x \mid Y)'$ so that $x \in kY$.

(K.II$_w$). For the first part, if $x \in X$ then $(x \mid X)'$ by S.III so that $x \in kX$. If $x \in k0$ then $(x \mid 0)'$ and hence $(0 \mid x)'$ by S.II and this is a contradiction to S.I.

(H.I$_w$). Using the notation of (2.1) we have to show that if $X \subset Y$ then $H(Y) \subset H(X)$. If $R \in H(Y)$ then $X \subset Y \subset R = kR$. From this it follows that $R \in H(X)$.

(H.II$_w$). For the first part it suffices to show that $kX \subset hX$ by virtue of K.II$_w$. But if $x \in kX$ it follows that $(x \mid R)'$ for any $R \in H(X)$, so that $x \in kR = R$. Thus x is contained in every element of $H(X)$ and thus $x \in hX$. For the second part, since $k0 = 0$ we have $0 \in H(0)$ and hence $h0 = 0$.

Before continuing with the proof of (2.5) we need the

(2.6) DEFINITION. *If X is a subset of an s_w-space and $X = kX$ then X is said to be k-closed.*

(2.7) THEOREM. *The common part of any aggregate of k-closed sets is k-closed.*

PROOF: Let X be the product of the family of sets $[X_\alpha]$ where $X_\alpha = kX_\alpha$. By K.I$_w$ we have $kX \subset kX_\alpha = X_\alpha$ so that $kX \subset X$. In consequence of K.II$_w$ we have $X = kX$.

Returning to the proof of (2.5): (H.III$_w$). We need only show that $hhX \subset hX$. By definition, hX is k-closed, i.e., we have $khX = hX$. But $hX \subset hX = khX$ and this qualifies hX for admission to the class $H(hX)$, and so $hhX \subset hX$.

(HK). We have proved thus far that

$$X \subset kX \subset hX = khX.$$

From the first inclusion and H.I$_w$ we get $hX \subset hkX$ and we know that $hkX \subset hhX = hX$ in consequence of K.II$_w$ and H.I$_w$. From this we have $hX = hkX$ and thus the proof of (2.5) is complete.

If the operator h (or k) is to be an analog of closure it should be additive. This requires that separation be additive. An s_w-space satisfying

AXIOM S.V. *Separation is additive*: $(X_1 \mid Y)\cdot(X_2 \mid Y) \rightarrow ((X_1 + X_2) \mid Y)$, will be called a *separation space* or an *s-space.*

Examples (2.2) and (2.3) are s-spaces while Example (2.4) is not. For if S is the Euclidean plane and X_1, Y and X_2 are respectively the points with coordinates $(-1, 0)$, $(0, 0)$ and $(1, 0)$ it is clear that S.V does not hold.

(2.8) THEOREM. *In an s-space* K.I$_w$ *and* H.I$_w$ *can be replaced by the stronger properties*

K.I $k(X + Y) = kX + kY$

H.I $h(X + Y) = hX + hY$.

PROOF: It follows readily from K.I$_w$ and H.I$_w$ that we have $kX + kY \subset k(X + Y)$ and $hX + hY \subset h(X + Y)$. If $x \in k(X + Y)$ this means that $(x \mid (X + Y))'$ and hence by S.II it follows that $((X + Y) \mid x)'$ and so by S.V either $(X \mid x)'$ or $(Y \mid x)'$. From this, using S.I, it is manifest that $x \in kX + kY$, proving K.I. To prove H.I we assume that $x \in h(X + Y)$ but is not a point of $hX + hY$. Then x is contained in every R in the class $H(X + Y)$ but there exist sets R_1 and R_2, respectively elements of $H(X)$ and $H(Y)$, which do not contain x. Now by K.I we have

$$R_1 + R_2 = kR_1 + kR_2 = k(R_1 + R_2) \supset X + Y$$

and x is not a point of $R_1 + R_2$. This a contradiction since $(R_1 + R_2) \in H(X + Y)$.

3. Topological Characterization of S-Spaces

We begin by stating

AXIOM S.VI. *If* $x \mid X$ *then* $x \mid kX$.

AXIOM S.VII. *If* x *and* y *are distinct points then* $x \mid y$.

An s-space satisfying the above two axioms will be called an s_k-*space.* Examples (2.2) and (2.3) are s_k-spaces. While axioms S.VI and S.VII hold in Example (2.4) it is not an s_k-space since it is not an s-space. The effect of these new postulates is shown by the

(3.1) THEOREM. *In an* s_k*-space the conditions* K.I–K.III *of* (3.3) *hold and* $kX = hX$.

PROOF: It is clear that S.VII and (2.5) imply K.II. In consequence of K.II$_w$ and K.I of (2.5) and (2.8) we need only show that $kkX \subset kX$. But this is immediate from S.VI. Further K.III implies $kX \in H(X)$ and so $hX \subset kX$. From HK of (2.5) we get equality here.

The fact that K.I–K.III of (3.3) hold in an s_k-space does not imply that $X \mid Y$ is equivalent to $X\cdot kY + Y\cdot kX = 0$.

(3.2) EXAMPLE. Let S be the unit interval with its usual topology and define "$(X \mid Y) \equiv (\bar{X}\cdot\bar{Y} = 0)$". The system $(S, \mid)$ is an s_k-space with $\bar{X} = hX = kX$. If P and Q denote respectively the sets $0 \leqq x < 1/2$ and $1/2 < x \leqq 1$ then $(P \mid Q)'$ but they satisfy the condition of Lennes-Hausdorff since they are top-

ologically separated. It is interesting to remark that S has no s-cut-points and indeed any subset R of S which is dense in S is s-connected (cf. §4).

This example indicates the need for the following

AXIOM S.VIII. *If for each $x \in X$ and each $y \in Y$ we have $(x \mid Y)$ and $(y \mid X)$ then $(X \mid Y)$.*

An s_k-space satisfying S.VIII will be called an *s_t-space.*

(3.3) CHARACTERIZATION THEOREM. *In an s_t-space we have*

K.I $k(X + Y) = kX + kY$.

K.II $kX = X$ *if X is null or a point.*

K.III $kkX = kX$.

R.I $(x \in kX) \equiv (x \mid X)'$.

R.II $(X \mid Y) \equiv (X \cdot kY + Y \cdot kX = 0)$.

Conversely, if the operator k satisfies K.I–K.III *and separation is defined by* R.II *then* S.I–S.VIII *and* R.I *hold.*

PROOF: It remains only to show, in consequence of (3.1) and the definitions, that R.II is valid. Suppose that the left member of R.II holds. If $X \cdot kY \neq 0$ (say) then there would be a point in X not separated from Y and hence $(X \mid Y)'$. If the right member holds and the left does not, then by S.VIII we can find a point $x \in X$ such that $(x \mid Y)'$. That is, $X \cdot kY \neq 0$, a contradiction. It is readily seen that the converse part of the theorem is valid.

It is to be noted that there exist spaces in which closure satisfies K.I, K.II$_w$ of (2.5), and K.III and in which separation (as defined by R.II) does not lead to R.I. This is simply because in such a space a point need not be a closed set, i.e., $kx \neq x$. Nevertheless we have

(3.4) THEOREM. *In an s-space, for any point x and any set X*

$$(x \mid X) \equiv (X \cdot kx + x \cdot kX = 0).$$

PROOF: The right member manifestly implies the left. Assume that the right member is false and first suppose that $x \cdot kX \neq 0$. Then by definition $(x \mid X)'$ and hence the left member fails to hold. Suppose only that $X \cdot kx \neq 0$. Then there is a $p \in X$ which is not separated from x. Consequently we have $(x \mid p)'$ and hence $x \in kp \subset kX$ in virtue of K.I$_w$ of (2.5). Again the left member fails to hold.

Topological spaces of the type (3.3) were studied extensively by Kuratowski [1] and others and are called T_1-spaces by Alexandroff-Hopf [1]. They may be described in terms of closed sets by saying that (i) the null set, the space and each point is closed (ii) the sum of two and the common part of any number of closed sets is again a closed set.

4. S-Connected Sets

In this section we assume that *S is an s-space* (Axioms S.I–S.V). Our purpose is to demonstrate that the "expected" theorems hold concerning "connected" sets.

(4.1) DEFINITION. *A subset of an s-space is said to be s-connected provided that it is not the sum of two non-null separated sets.*

(4.2) THEOREM. *If X is s-connected and contained in $Y + Z$, $Y \mid Z$, then $X \subset Y$ or $X \subset Z$.*

PROOF: If X intersects both Y and Z then $X = X \cdot Y + X \cdot Z$ and neither summand is null. By S.II, S.IV, and S.V it follows that $X \cdot Y \mid X \cdot Z$ and this is a contradiction to the definition of s-connectivity.

(4.3) THEOREM. *If $[X_\alpha]$ is a class of s-connected sets no one of which is separated from the s-connected set X, then the set $X_0 = X + \sum X_\alpha$ is s-connected.*

PROOF: If not, then $X_0 = Y + Z$ and $Y \mid Z$, neither being null. Then using (4.2) we may suppose that $X \subset Y$. Since $Z \cdot X_\alpha \neq 0$ for some α we have (again by (4.2)) that some X_α is contained in Z. By S.II, S.IV and S.V this implies that $X \mid X_\alpha$, a contradiction.

(4.4) THEOREM. *If X is a non-null s-connected subset of the set Y then there is a maximal s-connected subset X_0 of Y which contains X. The set X_0 is called the s-component of Y which contains X.*

PROOF: It is readily verified that the set $X_0 = X +$ [*all s-connected subsets of Y which are not separated from X*] is an s-connected subset of Y containing X and is not a proper subset of any other s-connected subset of Y having these properties. Hence X_0 is the desired set.

(4.5) THEOREM. *If Y is an s-connected subset of the s-connected set X then*

(i) *If $X - Y = P + Q$, $P \mid Q$, then $Y + P$ and $Y + Q$ are s-connected;*

(ii) *If Z is an s-component of $X - Y$ then $X - Z$ is s-connected.*

PROOF: (i). If we assume that $Y + P = M + N$ where $M \mid N$ and neither summand is null it follows by (4.2) that we may assume that Y is contained in N and hence by S.III that M is a subset of P. Hence, since,

$$X = (X - Y) + Y = P + Q + Y = M + (N + Q),$$

on applying S.IV and S.V it follows that $M \mid (N + Q)$, a contradiction to the assumption that X is s-connected.

(ii) Assume that $X - Z = M + N$ where $M \mid N$. Since Y does not intersect Z and Y is s-connected we may assume that Y is a subset of N. By (i) the set $M + Z$ is an s-connected subset of $X - Y$ and since Z is an s-component of this set and $M \cdot Z = 0$ we have $M = 0$.

A somewhat different form of the first part of (4.5) is as follows:

(4.6) THEOREM. *If P and $P \cdot Q$ are s-connected and $R \mid (P - (Q + R))$ then $P \cdot (Q + R)$ is s-connected.*

(4.7) THEOREM. *If $X + Y$ and $X \cdot Y$ are s-connected and $(X - Y) \mid (Y - X)$ then X and Y are s-connected.*

PROOF: We have

$$(X + Y) - X \cdot Y = (X - Y) + (Y - X) \quad \text{and} \quad X = X \cdot Y + (X - Y)$$

so that we may apply (4.5).

In (4.7) the condition $(X - Y) \mid (Y - X)$ replaces the classical (and un-

necessarily restrictive) "X and Y are closed in $X + Y$" which indeed is not applicable in an s-space. To show the efficacy of the condition it may be shown, for example, that *in order that a connected, locally connected metric space R be unicoherent it is necessary and sufficient that for any decomposition $R = P + Q$ where P and Q are connected and $(P - Q) \mid (Q - P)$ (topologically) we have $P \cdot Q$ connected.*

(4.8) THEOREM. *If X is s-connected and $X \subset X_0 \subset kX$ then X_0 is s-connected.*

PROOF. The set X_0 is of the form $X + \sum X_\alpha$ where each X_α is a point of $kX - X$. We then apply (4.3).

(4.9) THEOREM. *If X is s-connected so is hX.*

PROOF: From (4.8) we know that kX is s-connected. If $hX = M + N$ where $M \mid N$ then, because X is a subset of hX, we may assume that X is contained in M. Since $M \mid N$ it follows that $M \cdot kN = 0 = N \cdot kM$. But

$$M + N = hX = khX = k(M + N) = kM + kN$$

and thus $M = kM \supset X$ so that $hX \subset M$ and therefore $hX = M$. Hence $N = 0$ and we see that hX is s-connected.

If the set I is s-connected and contains the points a and b and no proper subset of I enjoys both of these properties then I is said to be *irreducibly s-connected between a and b.* It may be shown that if $x \in I$, $a \neq x \neq b$ then $I - x = I_1 + I_2$ where $a \in I_1$, $b \in I_2$, $I_1 \mid I_2$ and both I_1 and I_2 are connected. This enables one to order the elements of I (from a to b) by defining $p < q$ if there is an $x \in I$ giving the decomposition above with $p \in I_1$ and $q \in I_2$. We state without proof the[2]

THEOREM. *If I is an s-space satisfying Axiom* S.VIII *and I is irreducibly s-connected between a and b, then, if the elements of I are ordered as above, the Dedekind proposition holds.*

5. An Example

Let S be a space satisfying the first three axioms of Kuratowski (i.e., *closure* satisfies K.I–K.III of (3.3)) and let $G = [Z]$ be a family of closed sets of S. Throughout this section we *define* separation as follows: *The sets X and Y are separated* (i.e., $X \mid Y$) *provided that one of them is null or there exists an element Z of G such that*

$$(+) \quad S - Z = U + V, \qquad X \subset U, \qquad Y \subset V, \qquad \bar{U} \cdot V + U \cdot \bar{V} = 0.$$

(5.1) THEOREM. *The system $(S, |)$ is an s_w-space and for any subset P of S the sets kP and hP are closed and $\bar{P} \subset kP \subset hP$.*

PROOF: It is manifest that the $(S, |)$ is an s_w-space. We proceed to show that kP is closed. To this end it suffices to prove kP is the common part of a collection of sets $H = [U + Z]$ satisfying $(+)$. We define this collection as follows: For any x not in kP we know that $(+)$ holds with $P = X$ and $x = Y$ and we

[2] See Knaster and Kuratowski [1].

let H^* be the collection of all sets $U + Z$ thus obtained. If we let Q be the product of all the sets in H then $kP \subset Q$. The reverse inclusion is easily seen to be valid.

If the collection G contains only the null set then for each X in S the set kX is the sum of all the quasi-components of S which intersect $\bar{X}$. If G contains all closed sets and if further S has the property that for any two topologically separated subsets of S the condition $(+)$ holds, then the system $(S, |)$ is an s_1-space and $X_4 = kX = hX$. Further, condition R.II of (3.3) is valid so that separation and topological separation are the same. Thus varying the collection G enables us to get anything from an s_w-space to the full topology of S.

For the remainder of this section we suppose that the family G *is a multiplicative ideal in the lattice of closed subsets of* S, that is, if Z_1 and Z_2 are in G then so is $Z_1 + Z_2$, and if Z_0 is a closed subset of an element of G then $Z_0 \in G$.

(5.2) LEMMA. *If for* $i = 1, 2, \cdots n$ *we have*

$$(+)_i \qquad S - Z_i = U_i + V_i, \qquad X_i \subset U_i, \qquad Y_i \subset V_i,$$

$$\bar{U}_i \cdot V_i + U_i \cdot \bar{V}_i = 0, \qquad Z_i \in G,$$

then there is a $Z \in G$ *such that* $(+)$ *holds with* $X = X_1 \cdot X_2 \cdots \cdot X_n$ *and* $Y = Y_1 + Y_2 + \cdots + Y_n$.

PROOF: Since the general case follows by induction we restrict our attention to the case $n = 2$. Let $U = U_1 \cdot U_2$ and $V = V_1 + V_2$, and set $Z = S - (U + V)$. That $\bar{U} \cdot V + U \cdot \bar{V} = 0$ is immediate and clearly $X \subset U$ and $Y \subset V$. Further we have

$$\begin{aligned} Z = (S - U) \cdot (S - V) &= [(S - U_1) + (S - U_2)] \cdot [S - V] \\ &= [(Z_1 + V_1) + (Z_2 + V_2)] \cdot [S - V_1] \cdot [S - V_2] \\ &= (Z_1 - V_2) + (Z_2 - V_1). \end{aligned}$$

Since V_1 and V_2 are open it follows (since G is an ideal) that $Z \in G$.

(5.3) THEOREM. *The system* $(S, |)$ *is an s-space.*

PROOF: In virtue of (3.1) we need only prove that S.V holds. But if $X \mid Y_1$ and $X \mid Y_2$ then $(+)_i$ holds for $i = 1, 2$ and $X = X_1 = X_2$. Consequently $(+)$ holds with $Y = Y_1 + Y_2$.

The condition that G be an ideal is essential here, and even in this event it is possible to give examples of Peano spaces in which S.VI–S.VIII all fail to hold

(5.4) LEMMA. *If* P *is closed,* S *is compact metric, and for each* $p \in P$ *we have* $p \mid Q$, *then* $P \mid Q$.

PROOF. For each $p \in P$ condition $(+)$ holds with $X = Q$ and $p = Y$. Since B is compact and each V is open we can find an integer n such that $(+)_i$ holds for $i = 1, 2, \cdots n$, and $P \subset V_1 + V_2 + \cdots + V_n$. Thus $(+)$ holds with $X = Q$ and $Y = P$ and this completes the proof.

(5.5) THEOREM. *If S is compact metric and P is s-connected then kP and hP are connected.*

PROOF: If kP is not connected then we have $kP = M + N$ where M and N are disjoint closed sets, neither of which is void. There exists a closed subset F such that

$$S - F = A + B, \quad M \subset A, \quad N \subset B, \quad \bar{A} \cdot B + A \cdot \bar{B} = 0.$$

Since $F \cdot kP = 0$ we have $f \mid P$ for each $f \in F$ and thus by (5.4) we conclude that $F \mid P$. Thus (+) is valid with $X = P$ and $Y = F$. Since $P \subset A + B$ we may suppose that $P \cdot A \neq 0$, and thus $P \cdot (U - \bar{B}) \neq 0$ in virtue of the fact that $A \cdot \bar{B} = 0$. Further since $\bar{B} - B \subset F \subset V$ we have $U - B = U - \bar{B}$. It is clear that the sets $U - B$ and $V + B$ are topologically separated being disjoint and open, and moreover $S - Z = (U - B) + (V + B)$. But since P is s-connected and $P \cdot (U - B)$ is not null we have $P \subset U - B$. Further $N \cdot (V + B)$ is not null and so for some $q \in N \subset kP$ we have $q \mid P$, a contradiction. The proof that hP is connected is as follows: By (4.9) hP is s-connected and by (2.5) $khP = hP$ so that by what we have just shown the set hP is connected.

The following example has many interesting features.

(5.6) EXAMPLE. Let S be the unit interval with its usual topology, let s be the point 0, and let G consist of all closed subsets of the sequence $(1, 1/2, 1/3, \cdots, 1/n, \cdots 0)$. Notice that $s = hs = ks$ but that if X is a closed subset which does not contain the point s then $kX \neq s$ while $LX = s$. Here the only k-closed (h-closed) sets are the null set, the point s and S.

While the systems studied in this section have long been known in topology, going back to Poincaré (cf. Menger [1, 2], Moore [1]), I am not aware that any effort has been made to introduce a topology of separation, or to superimpose a topology of any sort on the system. The following proposition has a certain intrinsic interest.

(5.7) THEOREM. *The relationship of "not being separated" is a continuous function on the space of all closed sets of S in the sense that if $X_n \to X$ and $Y_n \to Y$, then $(X_n \mid Y_n)'$ implies $(X \mid Y)'$, S being assumed compact and metric.*

6. A Dynamical Example

In the present section we borrow the terminology of dynamics (cf. Hedlund [1]). Roughly speaking two sets will be "separated" if they are not "joined" by a motion. More exactly we consider a metric space S and a family $T = [t]$ of subsets of S (not necessarily closed). Further we suppose that through each point of S there passes at least one *motion* (= element of T). A subset P of S will be *invariant* provided that $P = \sum t, t \cdot P \neq 0$.

(6.1) DEFINITION. *The subsets X and Y of S will be separated* (i.e., $X \mid Y$) *provided that*

(i) *No element of T intersects both $\bar{X}$ and Y.*

(ii) *No element of T intersects both X and $\bar{Y}$.*

If the elements of T are points of S then separation and topological separation

are equivalent and the k-topology and topology of S agree. It may be seen that the following is valid.

(6.2) THEOREM. *The system $(S, |)$ is an s-space in which Axiom* S.VIII *holds. For any subset X of S we have*

$$kX = \sum t,\ t \cdot \bar{X} \neq 0.$$

In order that a set be k-closed it is necessary and sufficient that it be closed and invariant. The set hX is the smallest closed and invariant set containing X.

In general a set hx $(x \in S)$ will not be a minimal set (in the sense that it may contain a proper k-closed subset). It is readily seen that if the motions are connected sets then if P is s-connected sets kP and hP are also s-connected and connected.

7. Transformations and the Eilenberg-Whyburn Theorem

If A and B are s-spaces then the single-valued transformation $T(A) = B$ will be called *s-continuous* provided that

$$(X \mid Y) \equiv (T^{-1}(X) \mid T^{-1}(Y))$$

for each pair of subsets X, Y of B.

This definition is justified by the

(7.1) THEOREM. *If A and B are compact metric spaces and "separation" is interpreted to mean "topological separation" then $T(A) = B$ is continuous if and only if it is s-continuous.*

PROOF: If T is continuous it is readily seen that it is s-continuous. If T is not continuous we can find a sequence $x_n \to x$ in A such that $T(x_n)$ converges to a point $y \neq T(x)$. It follows that we may suppose that the sets $T(x)$ and $T(x_1) + T(x_2) + \cdots$ are separated. Hence the sets x and $x_1 + x_2 + \cdots$ will be separated in contradiction to the assumption that x is an accumulation point of the set $x_1 + x_2 + \cdots$.

The following proposition (while not immediately relevant to our purpose) has considerable intrinsic interest.

(7.2) THEOREM. *If $T(A) = B$ is continuous where A and B are compact metric then the following properties are equivalent ("separation" meaning "topological separation"):*

(i) *The image of an open set is open.*

(ii) *If $X \subset A$ and $Y \subset B$ then $(X \mid T^{-1}(Y)) \equiv (T(X) \mid Y)$.*

PROOF: Note first that (under our convention above) a set V is open if and only if $p \mid (B - V)$ for each $p \in V$. Hence if T satisfies (ii) it is to be shown that if U is open then $y \in T(U)$ implies $y \mid (B - T(U))$. Now if $x \in T^{-1}(y) \cdot U$ then x and $A - U$ are separated and hence the sets x and $A - T^{-1}T(U) = T^{-1}(B - T(U))$ are separated (in virtue of S.IV). It then follows by (ii) that $y \mid (B - T(U))$. We suppose now that T satisfies (i). To prove the leftward implication of (ii) we see that

$$(X \mid T^{-1}(Y))' \to (T^{-1}T(X) \mid T^{-1}(Y))' \to (T(X) \mid Y)'$$

by S.IV and (7.1). To prove the rightward implication we note that if (ii) holds then $T^{-1}(\bar{P}) = \overline{T^{-1}(P)}$ for each $P \subset B$, and $T(\bar{Q}) = \overline{T(Q)}$ for each $Q \subset A$. Hence

$$(X \mid T^{-1}(Y)) \rightarrow (\bar{X} \cdot T^{-1}(Y) + X \cdot T^{-1}(\bar{Y}) = 0)$$
$$\rightarrow \overline{(T(X)} \cdot Y + T(X) \cdot \bar{Y} = 0) \rightarrow (T(X) \mid Y).$$

We suppose from now on that A is an s-space and $T(A) = B$ is single-valued and varying hypotheses will be placed on T and B.

(7.3) *If B is any abstract set of elements and we define for each pair X, Y in B*

$$(X \mid Y) = (T^{-1}(X) \mid T^{-1}(Y)$$

then the system $(B, |)$ is an s-space and T is s-continuous

PROOF: To validate the second statement it remains only to prove the first. Since S.I–S.III are immediate we have to show that S.IV and S.V hold. But these are immediate in virtue of the identities $(P \subset Q) \equiv (T^{-1}(P) \subset T^{-1}(Q))$ and $T^{-1}(P + Q) = T^{-1}(P) + T^{-1}(Q)$, and the axioms.

(7.4) THEOREM. *If A' and B are s-spaces, $T_1(A) = A'$ and $T_2(A') = B$ are singled-valued, and $T = T_2T_1$ and T_1 are s-continuous, then T_2 is s-continuous.*

PROOF: If X and Y are subsets of B and $X \mid Y$, then we have $T^{-1}(X) \mid T^{-1}(Y)$. But $T^{-1}(X) = T_1^{-1}T_2^{-1}(X)$ and $T^{-1}(Y) = T_1^{-1}T_2^{-1}(Y)$ so that since T_1 is s-continuous we have $T_2^{-1}(X) \mid T_2^{-1}(Y)$. The reverse implication follows similarly so that T_2 is s-continuous.

(7.5) THEOREM. *If $T(A) = B$ is s-continuous (B an s-space) and $T^{-1}(b)$ is s-connected for each $b \in B$, then $T^{-1}(B_0)$ is s-connected for each s-connected subset B_0 of B.*

PROOF: Suppose that $T^{-1}(B_0) = P + Q$ where $P \mid Q$ and neither summand is void. Then for any $b \in B_0$ we have $T^{-1}(b)$ contained in either P or Q. Thus $P = T^{-1}T(P)$ and $Q = T^{-1}T(Q)$ so that we get $T(P) \mid T(Q)$. Since $B_0 = T(P) + T(Q)$ it follows that B_0 is not s-connected, a contradiction.

The following is a generalization of a theorem due to Eilenberg [1] and Whyburn [1]. It is to be noted that no use is made of upper semi-continuous decompositions in the proof of the theorem.

(7.6) THEOREM. *If A and B are s-spaces and $T(A) = B$ is s-continuous then there exists an s-space A' and s-continuous transformation $T_1(A) = A'$ and $T_2(A') = B$ such that $T = T_2T_1$ where*

(i) *for each $a' \in A'$ the set $T_1^{-1}(a')$ is connected and*

(ii) *for each $b \in B$ the set $T_2^{-1}(b)$ contains no s-connected set containing more than one point.*

PROOF: Let $A' = [a']$ be the class consisting of all s-components of all sets $T^{-1}(b)$, $b \in B$. For each $a \in A$ let $T_1(a) = a'$ be the s-component of $T^{-1}T(a)$ which contains a. For $X', Y' \subset A'$ define $X' \mid Y'$ to mean that $T_1^{-1}(X') \mid T_1^{-1}(Y')$. Then by (7.3) T_1 is s-continuous and A' is an s-space. By (7.4) $T_2 \equiv T_1T^{-1}$ is s-continuous. If X' is an s-connected subset of $T_2^{-1}(b)$, $b \in B$,

then by (7.5) we know that $T_1^{-1}(X')$ is s-connected and is contained in $T_1^{-1}T_2^{-1}(b) = T^{-1}(b)$. Thus $T_1^{-1}(X')$ is in a single component of $T^{-1}(b)$ and thus X' is a single point of A'.

THE UNIVERSITY OF VIRGINIA

BIBLIOGRAPHY

Alexandroff-Hopf.....[1], *Topologie* I, Berlin.
Eilenberg.....[1], Fund. Math. 22(1934), p. 292.
Hedlund.....[1], Bull. Am. Math. Soc. 45(1939), p. 241.
Knaster and Kuratowski.....[1], Fund. Math. 2(1921), p. 206.
Kuratowski.....[1], *Topologie* I, Warsaw.
Menger.....[1], *Dimensionstheorie*, Berlin.
Menger.....[2], *Kurventheorie*, Berlin.
Moore.....[1], *Foundations of Point Set Theory*, New York.
Whyburn.....[1], Am. Journ. Math. 56(1934), p. 370.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

TENSOR DECOMPOSITION WITH APPLICATIONS TO THE CONTACT AND COMPLEX GROUPS

BY T. C. DOYLE

(Received September 12, 1939)

INTRODUCTION

The original purpose of this paper was to apply tensor methods to a study of the invariant theory of homogeneous contact transformations. L. P. Eisenhart and M. S. Knebelman (1) have shown how a vector Λ of $2q$ components in the phase space of $2q$ coordinates x^i and p_i may be decomposed into two geometric objects λ and μ of q components each. The transformation laws of λ and μ were such that whereas the new components $\bar{\lambda}$ of one set were linear combinations of the corresponding old components λ, the new components $\bar{\mu}$ were linearly dependent on both sets λ and μ. Because of this transformation behavior the decomposition just cited will be described as *partial*. In this paper we shall arrive at a complete decomposition in the sense that the $\bar{\lambda}$'s will be defined in terms of the λ's only and the $\bar{\mu}$'s in terms of the μ's. The underlying formalism of the Eisenhart-Knebelman paper has been employed here and the decomposition at which we arrive agrees, when restricted to extended point transformations, with their similarly restricted decomposition.

Our decomposition has been accomplished by introducing the notion of two families of complementary subspaces of the phase space and of projection tensors in these subspaces. Since the method is independent of any implication of a contact transformation on the coordinates it will first be developed in its full generality and the decomposition under the contact group will be presented as a special application of the general theory. As additional applications we shall give two decompositions of tensors under the complex group, one being real and the other complex. In section 16 the doubly homogeneous contact group of Schouten (2) is derived independently and a complete decomposition for this group in terms of an affine connection is given.

1. THE UNDERLYING SYSTEMS OF DIFFERENTIAL EQUATIONS

Two systems of partial differential equations of the solved type

$$X_\alpha f = L^\varrho{}_\alpha \partial_\varrho f = \partial_\alpha f + L^\tau{}_\alpha(x^A)\,\partial_\tau f = 0 \tag{1.1}$$

$$X_i f = L^\varrho{}_i \partial_\varrho f = \partial_i f + L^\rho{}_i(x^A)\,\partial_\rho f = 0 \tag{1.2}$$

for which $|L^A{}_B| = \begin{vmatrix} \delta^\alpha{}_\beta & L^\alpha{}_j \\ L^i{}_\beta & \delta^i{}_j \end{vmatrix} \neq 0$ and $\partial_A f = \dfrac{\partial f}{\partial x^A}$ and wherein the indices have the ranges $(\alpha) = (1, \cdots, q)$, $(i) = (q+1, \cdots, N)$, and $(A) = (1, \cdots, N)$ will be called *complementary*.

Associated with the systems $X_i f = 0$ and $X_\alpha f = 0$ are their dual Pfaffian systems $P^\alpha = N^\alpha{}_Q\, dx^Q = dx^\alpha - L^\alpha{}_r\, dx^r = 0$ and $P^i = N^i{}_Q\, dx^Q = dx^i - L^i{}_\rho\, dx^\rho = 0$ with the matrix $\| N^A{}_B \| = \begin{Vmatrix} \delta^\alpha{}_\beta & -L^\alpha{}_j \\ -L^i{}_\beta & \delta^i{}_j \end{Vmatrix}$. From these definitions, if $f(x)$ is a solution of $X_i f = 0$ its differential $df = \partial_\alpha f(dx^\alpha - L^\alpha{}_r\, dx^r)$ will vanish for any increment dx^A satisfying $P^\alpha = 0$. Conversely, if $f(x)$ is an integral of $P^\alpha = 0$ then $df = (\partial_i f + L^\rho{}_i \partial_\rho f)\, dx^i$ must vanish identically in dx^i and $f(x)$ is a solution of $X_i f = 0$.

We shall confine our attention to (1.1) in making a few general remarks. Assuming for the moment complete integrability, there exist $N - q$ principal solutions (3) $\varphi^i(x^A)$ satisfying the initial conditions $\varphi^i(x_0^\alpha, x^j) = x^i$ within a suitably small neighborhood of a point $(x_0{}^A)$ for which $L^i{}_\alpha(x^A)$ are regular and the most general system of solutions is given by any $N - q$ independent functions $f^i(\varphi^j)$ of the principal solutions. Since $\partial_j \varphi^i \,|_{x^\alpha = x_0^\alpha} = \delta^i{}_j$ it follows that the determinant $| \partial_j f^i |$ does not vanish identically. A general system of solutions $f^i(x)$ determines an ∞^{N-q} family of q-dimensional subspaces $S^q(c^i)$ as defined by the equations $f^i(x) = c^i$. A parametric representation may be obtained by solving these equations to give $x^i = \psi^i(x^\alpha, c^j)$, the complete parametric system for $S^q(c^i)$ being

$$x^\alpha = x^\alpha \qquad x^i = \psi^i(x^\alpha, c^j). \tag{1.3}$$

Now the functions $F^i = x^i - \psi^i(x^\alpha, c^j)$ when equated to zero represent the same family of subspaces $S^q(c^i)$ as defined by $f^i(x) = c^i$ and hence must satisfy the system (1.1) for all values of x^A lying in $S^q(c^i)$, that is,

$$\partial_\alpha \psi^i = L^i{}_\alpha(x^\beta, \psi^j(x^\beta, c^k)). \tag{1.4}$$

Suppose $S^q(c^i)$ to be parameterized in a general manner by the equations $x^A = x^A(u^\alpha)$. Then the q invariants $\lambda_\alpha = \Lambda_Q \dfrac{\partial x^Q}{\partial u^\alpha}$ are the scalar products of Λ_A with the q tangent vectors $\dfrac{\partial x^A}{\partial u^\alpha}$ of $S^q(c^i)$. Under parameter transformations the λ's constitute a q-component vector of $S^q(c^i)$ which will be called the projection of Λ_A on $S^q(c^i)$. A distinguishing feature of the parameterization (1.3) may be emphasized by writing it in the form

$$x^\alpha = u^\alpha \qquad x^i = \psi^i(u^\alpha, c^j) \tag{1.5}$$

and transforming the x's by $\bar{x}^A = f^A(x)$ so that $\bar{x}^\alpha = f^\alpha(x^\beta, \psi^i(x^\beta, c^j))$. If we insist on invariance of form for the first set of (1.5), namely $\bar{x}^\alpha = \bar{u}^\alpha$, it may be seen that a transformation $x^A \to \bar{x}^A$ induces the parameter transformation $u \to \bar{u}$ given by $\bar{u}^\alpha = f^\alpha(u^\beta, \psi^i(u^\beta, c^j))$. The condition for non-singularity is $\left| \dfrac{\partial \bar{u}^\alpha}{\partial u^\beta} \right| = \left| \dfrac{\partial f^\alpha}{\partial u^\beta} + \dfrac{\partial f^\alpha}{\partial \psi^r} \dfrac{\partial \psi^r}{\partial u^\beta} \right| \neq 0$. Reverting to our first notation $x^\alpha = u^\alpha$ and applying (1.4) this condition on the transformation $x^A \to \bar{x}^A$ takes the form $| X_\beta f^\alpha | = | \partial_\beta f^\alpha + L^r{}_\beta \partial_r f^\alpha | \neq 0$ for points (x^A) of $S^q(c^i)$.

It has just been shown how every coordinate transformation $x^A \to \bar{x}^A$ satisfying $|X_\beta f^\alpha| \neq 0$ induces an isomorphic non-singular parameter transformation $u^\alpha \to \bar{u}^\alpha$. Thus the coordinate scalars $\lambda_\alpha = \Lambda_Q \dfrac{\partial x^Q}{\partial u^\alpha}$ will transform as parameter vectors under a coordinate transformation $x \to \bar{x}$ and because of the covariantive parameterization $x^\alpha = u^\alpha$ these parameter vectors will have the semblance of q-component coordinate vectors. The components λ_α are $\lambda_\alpha(x^A) = \Lambda_\alpha + \Lambda_r L^r{}_\alpha$ where the x's are confined to a subspace $S^q(c^i)$. But since any point (x^A) is a point of some such subspace, the quantities λ_α constitute a q-component *factor* field of the N-component vector field Λ_A. Analogous conclusions may be drawn from the complementary system (1.2) leading to the complete decomposition $\lambda_\alpha = \Lambda_Q L^Q{}_\alpha = \Lambda_\alpha + \Lambda_r L^r{}_\alpha$ and $\lambda_i = \Lambda_Q L^Q{}_i = \Lambda_i + \Lambda_\rho L^\rho{}_i$ of Λ_A into a q and $N - q$ component factor field respectively. The restriction $|X_j f^i| = |\partial_j f^i + L^\rho{}_j \partial_\rho f^i| \neq 0$ on the coordinate transformation $\bar{x}^A = f^A(x)$ analogous to $|X_\beta f^\alpha| \neq 0$ is encountered. To determine the transformation character of the factor fields we shall investigate the behavior of the basic systems of partial differential equations under a change of coordinates.

2. The Transformation Character of the Basic Systems

Let the most general coordinate transformation $x \to \bar{x}$ be sought which will preserve the solved form of the basic systems. That is, we stipulate the general laws

$$\begin{aligned} \bar{X}_\beta f &= T_\beta{}^\rho X_\rho f & \bar{X}_j f &= T_j{}^r X_r f \\ P^\alpha &= \bar{P}^\rho T_\rho{}^\alpha & P^i &= \bar{P}^r T_r{}^i \end{aligned} \tag{2.1}$$

wherein $|T_\beta{}^\alpha| \neq 0$ and $|T_j{}^i| \neq 0$ so that $X_\beta f = 0$ will imply $\bar{X}_\beta f = \bar{\partial}_\beta f + \bar{L}^r{}_\beta \bar{\partial}_r f = 0$ etc. If both sides of these equations be written as linear forms in $\bar{\partial}_A f$ and $d\bar{x}^A$ and the coefficients be identified there will result

$$\begin{aligned} \bar{L}^i{}_\rho X_\beta \bar{x}^\rho &= X_\beta \bar{x}^i & \bar{L}^\alpha{}_r X_j \bar{x}^r &= X_j \bar{x}^\alpha \\ T_\beta{}^\rho X_\rho \bar{x}^\alpha &= \delta^\alpha{}_\beta & T_j{}^r X_r \bar{x}^i &= \delta^i{}_j \end{aligned} \tag{2.2}$$

$$\begin{aligned} \bar{L}^r{}_\beta \bar{X}_r x^i &= -\bar{\partial}_\beta x^i + L^i{}_\rho \bar{\partial}_\beta x^\rho & \bar{L}^\rho{}_j \bar{X}_\rho x^\alpha &= -\bar{\partial}_j x^\alpha + L^\alpha{}_r \bar{\partial}_j x^r \\ \bar{L}^r{}_\beta \bar{\partial}_r x^\alpha &= -L^\alpha{}_r \bar{\partial}_\beta x^r & \bar{L}^\rho{}_j \bar{\partial}_\rho x^i &= -L^i{}_\rho \bar{\partial}_j x^\rho. \end{aligned} \tag{2.3}$$

The first line of (2.3) reduces to the inverse of the corresponding line in (2.2) when one applies the second line of (2.3). We impose the restrictions

$$(R_{2.4}) \qquad |X_\beta \bar{x}^\alpha| \neq 0 \qquad |X_j \bar{x}^i| \neq 0$$

identical with those of the preceding section on the transformations $\bar{x}^A = \bar{x}^A(x)$ so that $\bar{L}^i{}_\beta$ and $\bar{L}^\alpha{}_j$ will be determined uniquely by (2.2). The second set of (2.3) may be regarded as identities which obviously hold for the identical transformation $\bar{x}^A = x^A$. Transformations $x \to \bar{x}$ satisfying $(R_{2.4})$ will be called

admissible and henceforth only such transformations will be considered. Equations (2.2) are adequate to establish the chain formulas of differentiation

$$X_\beta \bar{x}^\alpha = \bar{X}_\rho \bar{x}^\alpha X_\beta \bar{x}^\rho \qquad X_j \bar{x}^i = \bar{X}_r \bar{x}^i X_j \bar{x}^r \tag{2.5}$$

giving the conjugate relationships

$$\delta^\alpha{}_\beta = \bar{X}_\rho x^\alpha X_\beta \bar{x}^\rho \qquad \delta^i{}_j = \bar{X}_r x^i X_j \bar{x}^r \tag{2.6}$$

for the transformation $\bar{x}^A = x^A$. Comparison with the second line of (2.2) gives $T_\beta{}^\alpha = \bar{X}_\beta x^\alpha$ and $T_j{}^i = \bar{X}_j x^i$ and from (2.5) the transitivity of the transformations $L \to \bar{L}$ follows readily. The transformation equations (2.1) of the basic systems are thus

$$\begin{aligned} \bar{X}_\beta f &= X_\rho f \bar{X}_\beta x^\rho \qquad \bar{X}_j f = X_r f \bar{X}_j x^r \\ P^\alpha &= \bar{P}^\rho \bar{X}_\rho x^\alpha \qquad P^i = \bar{P}^r \bar{X}_r x^i. \end{aligned} \tag{2.7}$$

3. The Group $\Sigma \frown \Sigma^*$ of Admissible Transformations

Although the transformations $x \to \bar{x}$ have been defined as changes of coordinates, it will be convenient in this section to regard them as point transformations in order to facilitate the description of the transitive properties of certain objects transforming under them. We shall now establish the group properties of the transformations of the set Σ: $\bar{x}^A = \sigma^A(x)$ characterized by $| X_\beta \bar{x}^\alpha | \neq 0$. We first observe that the identical transformation $\bar{x}^A = x^A$ is in Σ. Secondly, if we first transform from the point (x) to the point $(\bar{x})$ by a transformation of Σ, thereby determining at $(\bar{x})$ a unique set of functions $\bar{L}^i{}_\beta(\bar{x})$, and then transform to a third point $(\bar{\bar{x}})$ by any transformation $\bar{\bar{x}}^A = \bar{\sigma}^A(\bar{x})$ of a set which we shall call $\bar{\Sigma}$, where the elements of $\bar{\Sigma}$ are characterized by the property that $| \bar{X}_\beta \bar{\bar{x}}^\alpha | \neq 0$, it is always possible to transform directly from (x) to $(\bar{\bar{x}})$ by an element of Σ, that is, by a transformation $\bar{\bar{x}}^A = \bar{\bar{x}}^A(x)$ for which $| X_\beta \bar{\bar{x}}^\alpha | \neq 0$; for it is merely necessary to form the determinant of both sides of (2.5) and observe that $| X_\beta \bar{\bar{x}}^\alpha |$ is the product of the two non-zero determinants $| X_\beta \bar{x}^\alpha |$ and $| \bar{X}_\beta \bar{\bar{x}}^\alpha |$ and so must be likewise non-zero itself. It remains to show that if $x \to \bar{x}$ is an element of Σ the inverse transformation $x \leftarrow \bar{x}$ is an element of $\bar{\Sigma}$, that is, we must prove that if $| X_\beta \bar{x}^\alpha | \neq 0$ then $| \bar{X}_\beta x^\alpha | \neq 0$. But this follows immediately from (2.6), $1 = | \bar{X}_\rho x^\alpha | \cdot | X_\beta \bar{x}^\rho |$. Since it has now been proved that the most general transformation $\bar{x} \to \bar{\bar{x}}$ with which we shall be concerned is compounded from an inverse transformation of the set Σ followed by a direct transformation of Σ we have the

Theorem 3.1: *Corresponding to any choice of the functions $L^i{}_\beta(x)$ in some arbitrary but definite coordinate system x, the subset Σ of the general group G of analytic transformations $\bar{x}^A = \bar{x}^A(x)$ for which $| X_\beta \bar{x}^\alpha | \neq 0$ when taken in conjunction with all the inverse transformations of Σ constitutes a subgroup of G.*

An analogous theorem holds for the subset Σ^* of G for which $| X_j \bar{x}^i | \neq 0$. It follows that the transformations of the intersection set $\Sigma \frown \Sigma^*$ likewise form a subgroup of G.

4. Factorization of Any Tensor

The invariance of the independence condition $|L^A{}_B| \neq 0$ will now be established by proving that after an admissible transformation $x \to \bar{x}$ characterized by $|\partial_B \bar{x}^A| \neq 0$ and $|\Delta_B \bar{x}^A| = \begin{vmatrix} X_\beta \bar{x}^\alpha & 0 \\ 0 & X_j \bar{x}^i \end{vmatrix} \neq 0$ the new determinant $|\bar{L}^A{}_B|$ will be likewise non-vanishing. The transformation laws (2.7) for the double system $X_B f = 0$ become $L^Q{}_B \partial_Q f = \bar{L}^A{}_Q \Delta_B \bar{x}^Q \bar{\partial}_A f$ and the conditions that these hold for arbitrary f are $L^Q{}_B \partial_Q \bar{x}^A = \bar{L}^A{}_Q \Delta_B \bar{x}^Q$. On forming the determinant of both sides it is apparent that $|\bar{L}^A{}_B| \neq 0$ is a consequence of $|L^A{}_B| \neq 0$, $|\partial_B \bar{x}^A| \neq 0$ and $|\Delta_B \bar{x}^A| \neq 0$ as was to be proved.

We shall call the matrix $||L^A{}_B||$ the *factor frame* of S^N, a name suggested by the definitions $\lambda_A = \Lambda_Q L^Q{}_A$ of the factor components λ_α and λ_i corresponding to Λ_A. Thus the left members $X_B f$ of our underlying equations are factor components of the gradient $\partial_A f$. In a new coordinate system $\bar{x}$ the factor components of an arbitrary vector $\bar{\Lambda}_B$ are defined by $\bar{\lambda}_B = \bar{\Lambda}_Q \bar{L}^Q{}_B$. For the case $\Lambda_A = \partial_A f$ we know from (2.7) that $\lambda_B = \bar{\lambda}_Q \Delta_B \bar{x}^Q$. To verify that these hold for an arbitrary vector $\bar{\Lambda}_A$ we rewrite them as $(\bar{L}^A{}_Q \Delta_B \bar{x}^Q - L^Q{}_B \partial_Q \bar{x}^A)\bar{\Lambda}_A = 0$ and note that the coefficients of $\bar{\Lambda}_A$ vanish as a consequence of the definitions of $\Delta_B \bar{x}^A$ and of the transformation laws (2.2) of the L's.

Up to this point we have considered only a factorization for a covariant vector Λ_A with factor components $\lambda_A = \Lambda_Q L^Q{}_A$ transforming according to the laws $\bar{\lambda}_\alpha = \lambda_\rho \bar{X}_\alpha x^\rho$ and $\bar{\lambda}_i = \lambda_r \bar{X}_i x^r$. Because of the condition $|L^A{}_B| \neq 0$ there exists a matrix $||M^A{}_B||$ conjugate to $||L^A{}_B||$ satisfying $M^A{}_Q L^Q{}_B = \delta^A{}_B$. As a basis for defining the factor components λ^A of a contravariant vector Λ^A we shall insist that $\lambda^Q \lambda_Q = \Lambda^Q \Lambda_Q$ hold for Λ^A and Λ_A arbitrary. This leads to the definitions $\lambda^A = M^A{}_Q \Lambda^Q$. From the four sets of relations

$$(4.1) \qquad \begin{aligned} M^\alpha{}_\beta + M^\alpha{}_r L^r{}_\beta &= \delta^\alpha{}_\beta & M^\alpha{}_j + M^\alpha{}_\rho L^\rho{}_j &= 0 \\ M^i{}_j + M^i{}_\rho L^\rho{}_j &= \delta^i{}_j & M^i{}_\beta + M^i{}_r L^r{}_\beta &= 0 \end{aligned}$$

resulting from $M^A{}_Q L^Q{}_B = \delta^A{}_B$ these frame components are

$$(4.2) \qquad \lambda^\alpha = M^\alpha{}_\beta(\Lambda^\beta - L^\beta{}_r \Lambda^r) \qquad \lambda^i = M^i{}_j(\Lambda^j - L^j{}_\rho \Lambda^\rho)$$

where the coefficients $M^\alpha{}_\beta$ and $M^i{}_j$ are defined by the conjugate relations

$$(4.3) \qquad M^\alpha{}_\rho(\delta^\rho{}_\beta - L^\rho{}_r L^r{}_\beta) = \delta^\alpha{}_\beta \qquad M^i{}_r(\delta^r{}_j - L^r{}_\sigma L^\sigma{}_j) = \delta^i{}_j .$$

It will now be shown how a factorization of an arbitrary covariant vector will induce a corresponding factorization for an arbitrary contravariant vector. For if we define factor components λ^A of Λ^A in two coordinate systems x and $\bar{x}$ by $\lambda^A = M^A{}_Q \Lambda^Q$ and $\bar{\lambda}^A = \bar{M}^A{}_Q \bar{\Lambda}^Q$ the two sides of $\bar{\Lambda}^A \bar{\Lambda}_A = \Lambda^Q \Lambda_Q$ will be equal respectively to the two sides of $\bar{\lambda}^A \bar{\lambda}_A = \lambda^Q \lambda_Q$, or $\bar{\lambda}^\alpha \bar{\lambda}_\alpha + \bar{\lambda}^i \bar{\lambda}_i = \lambda^\rho(\bar{\lambda}_\alpha X_\rho \bar{x}^\alpha) + \lambda^r(\bar{\lambda}_i X_r \bar{x}^i)$ by virtue of our definitions of λ_A and $\bar{\lambda}_A$. By insisting that these be identities in $\bar{\lambda}_A$ we obtain the desired laws $\bar{\lambda}^\alpha = \lambda^\rho X_\rho \bar{x}^\alpha$ and $\bar{\lambda}^i = \lambda^r X_r \bar{x}^i$. It involves but a slight extension of this demonstration to define factor com-

ponents $\lambda^A{}_B$ of a simple mixed tensor $\Lambda^A{}_B$ by $\lambda^A{}_B = M^A{}_Q \Lambda^Q{}_R L^R{}_B$ and observe that they will transform by $\bar{\lambda}^\alpha{}_\beta = \lambda^\rho{}_\sigma X_\rho \bar{x}^\alpha \bar{X}_\beta x^\sigma$, $\bar{\lambda}^i{}_j = \lambda^r{}_s X_r \bar{x}^i \bar{X}_j x^s$, $\bar{\lambda}^\alpha{}_j = \lambda^\rho{}_r X_\rho \bar{x}^\alpha \bar{X}_j x^r$ and $\bar{\lambda}^i{}_\beta = \lambda^r{}_\rho X_r \bar{x}^i \bar{X}_\beta x^\rho$. The generalization to tensors of arbitrary covariant and contravariant rank is obvious.

Let us now return to equations (4.3). Defining $t^\alpha{}_\beta$ and $t^i{}_j$ by $t^\alpha{}_\beta = \delta^\alpha{}_\beta - L^\alpha{}_r L^r{}_\beta$ and $t^i{}_j = \delta^i{}_j - L^i{}_\rho L^\rho{}_j$ the equations (4.3) are $M^\alpha{}_\rho t^\rho{}_\beta = \delta^\alpha{}_\beta$ and $M^i{}_r t^r{}_j = \delta^i{}_j$. Their form suggests that $t^\alpha{}_\beta$, $t^i{}_j$, $M^\alpha{}_\beta$, and $M^i{}_j$ are factor tensors and such is in fact the case. To establish the tensor character of $t^\alpha{}_\beta$ it is sufficient to observe that $(\bar{L}^\alpha{}_r \bar{L}^r{}_\rho) X_\beta \bar{x}^\rho = (L^\rho{}_\sigma L^\sigma{}_\beta) X_\rho \bar{x}^\alpha$ follows from (2.2) and (2.3). By use of the t's the transformation laws (2.2) of the L's may be written in the form

$$(4.4) \qquad \bar{L}^i{}_\rho X_\beta \bar{x}^\rho = L^r{}_\beta X_r \bar{x}^i + t^\rho{}_\beta \partial_\rho \bar{x}^i \qquad \bar{L}^\alpha{}_r X_j \bar{x}^r = L^\rho{}_j X_\rho \bar{x}^\alpha + t^\rho{}_j \partial_r \bar{x}^\alpha.$$

It is now apparent that corresponding to any contravariant vector Λ^A there are the two sets of factor vectors

$$\lambda^{*\alpha} = \Lambda^\alpha - L^\alpha{}_r \Lambda^r \qquad \lambda^{*i} = \Lambda^i - L^i{}_\rho \Lambda^\rho$$

$$\lambda^\alpha = M^\alpha{}_\rho \lambda^{*\rho} \qquad \lambda^i = M^i{}_r \lambda^{*r}.$$

These two sets are identical when and only when the tensors $L^\alpha{}_r L^r{}_\beta$ and $L^i{}_\rho L^\rho{}_j$ vanish.

Referring to the definitions in section 1 of the coefficients $N^A{}_B$ of the fundamental Pfaffians, it may be seen that $\| N^A{}_Q L^Q{}_B \| = \begin{Vmatrix} t^\alpha{}_\beta & 0 \\ 0 & t^i{}_j \end{Vmatrix}$ and therefore $| N^A{}_Q | \cdot | L^Q{}_B | = | t^\alpha{}_\beta | \cdot | t^i{}_j |$. But since $| N^A{}_B | = | L^A{}_B |$, as may be seen by multiplying the last $N - q$ columns (index j) and the last $N - q$ rows (index i) of $| N^A{}_B |$ by -1, we have $| L^A{}_B |^2 = | t^\alpha{}_\beta | \cdot | t^i{}_j |$ and in new coordinates $\bar{x}$, $| \bar{L}^A{}_B |^2 = | \bar{t}^\alpha{}_\beta | \cdot | \bar{t}^i{}_j |$. Because the t's are simple mixed factor tensors, $| \bar{t}^\alpha{}_\beta | = | t^\alpha{}_\beta |$ and $| \bar{t}^i{}_j | = | t^i{}_j |$, and hence $| \bar{L}^A{}_B |^2 = | L^A{}_B |^2$. From continuity considerations under an infinitesimal transformation we finally have $| \bar{L}^A{}_B | = | L^A{}_B |$. Hence from $L^Q{}_B \partial_Q \bar{x}^A = \bar{L}^A{}_Q \Delta_B \bar{x}^Q$ there follows the relation $| \partial_B \bar{x}^A | = | X_\beta \bar{x}^\alpha | \cdot | X_j \bar{x}^i |$.

5. Covariant Differentiation and the Commutator Symbols

We now assume the existence in S^N of an affine connection $L^A{}_{BC}(x)$ in terms of which we define the covariant derivatives of arbitrary vectors by $\Lambda_{C,B} = \partial_B \Lambda_C - \Lambda_Q L^Q{}_{BC}$ and $\Lambda^A{}_{,B} = \partial_B \Lambda^A + \Lambda^Q L^A{}_{BQ}$. Similarly for the factor components of these vectors we define factor covariant derivatives by $\lambda_{C,B} = X_B \lambda_C - \lambda_Q l^Q{}_{BC}$ and $\lambda^A{}_{,B} = X_B \lambda^A + \lambda^Q l^A{}_{BQ}$, where the l's are functions to be called factor components of connection, undetermined as yet, such that these latter quantities $\lambda_{C,B}$ and $\lambda^A{}_{,B}$ will be factor components of the former tensors $\Lambda_{C,B}$ and $\Lambda^A{}_{,B}$ respectively; $\lambda_{C,B} = \Lambda_{R,Q} L^R{}_C L^Q{}_B$ and $\lambda^Q{}_{,B} L^A{}_Q = \Lambda^A{}_{,Q} L^Q{}_B$. These conditions determine

$$(5.1) \qquad l^Q{}_{BC} L^A{}_Q = L^A{}_{QR} L^Q{}_B L^R{}_C + L^Q{}_B \partial_Q L^A{}_C$$

as the relations that must exist between $l^A{}_{BC}$ and $L^A{}_{BC}$. In a new coordinate system $\bar{x}$ we have components of affine connection given by $\bar{L}^Q{}_{BC}\bar{\partial}_Q x^A = L^A{}_{QR}\bar{\partial}_B x^Q \bar{\partial}_C x^R + \bar{\partial}^2_{BC} x^A$ and these induce the transformation law

$$l^Q{}_{BC}\bar{\Delta}_Q x^A = l^A{}_{QR}\bar{\Delta}_B x^Q \bar{\Delta}_C x^R + \bar{X}_B \bar{\Delta}_C x^A \tag{5.2}$$

according to which

$$\begin{aligned} l^\rho{}_{\beta\gamma}\bar{X}_\rho x^\alpha &= l^\alpha{}_{\rho\sigma}\bar{X}_\beta x^\rho \bar{X}_\gamma x^\sigma + \bar{X}_\beta \bar{X}_\gamma x^\alpha \quad l^\rho{}_{j\gamma}\bar{X}_\rho x^\alpha = l^\alpha{}_{r\sigma}\bar{X}_j x^r \bar{X}_\gamma x^\sigma + \bar{X}_j \bar{X}_\gamma x^\alpha \\ l^r{}_{jk}\bar{X}_r x^i &= l^i{}_{rs}\bar{X}_j x^r \bar{X}_k x^s + \bar{X}_j \bar{X}_k x^i \qquad l^r{}_{\beta k}\bar{X}_r x^i = l^i{}_{\rho s}\bar{X}_\beta x^s \bar{X}_k x^s + \bar{X}_\beta \bar{X}_k x^i \end{aligned} \tag{5.3}$$

while $l^i{}_{\beta\gamma}$, $l^\alpha{}_{jk}$, $l^i{}_{j\gamma}$ and $l^\alpha{}_{\beta k}$ transform as factor tensors.

The skew-symmetric portion $\Omega^A{}_{BC}$ of the affine connection $L^A{}_{BC} = \Gamma^A{}_{BC} + \Omega^A{}_{BC}$ is a tensor with factor components $\omega^A{}_{BC}$ which may be computed from (5.1),

$$\tfrac{1}{2}(L^A{}_{QR} - L^A{}_{RQ})L^Q{}_B L^R{}_C = \tfrac{1}{2}(l^Q{}_{BC} - l^Q{}_{CB})L^A{}_Q - \tfrac{1}{2}I^A{}_{BC} \tag{5.4}$$

where $I^A{}_{BC} = L^Q{}_B\partial_Q L^A{}_C - L^Q{}_C\partial_Q L^A{}_B$ so that $I^\alpha{}_{\beta\gamma} = I^i{}_{jk} = 0$. The commutator $(X_B X_C)f$ is then $(X_B X_C)f = I^Q{}_{BC}\partial_Q f = M^A{}_Q I^Q{}_{BC} L^R{}_A \partial_R f = \lambda^A{}_{BC} X_A f$ where $\lambda^A{}_{BC} = M^A{}_Q I^Q{}_{BC}$. The quantities $\lambda^A{}_{BC}$ will be called *commutator symbols* for the operator X_B and in terms of them we have by (5.4), $\Omega^A{}_{QR}L^Q{}_B L^R{}_C = \tfrac{1}{2}[(l^Q{}_{BC} - l^Q{}_{CB}) - \lambda^Q{}_{BC}]L^A{}_Q$, and hence the factor components $\omega^A{}_{BC}$ of $\Omega^A{}_{BC}$ are

$$\omega^A{}_{BC} = \tfrac{1}{2}[(l^A{}_{BC} - l^A{}_{CB}) - \lambda^A{}_{BC}]. \tag{5.5}$$

6. Integrability Conditions and Normal Coordinates

The system $X_\beta f = 0$ will be completely integrable if and only if the commutators $(X_\beta X_\gamma)f = \lambda^\rho{}_{\beta\gamma}X_\rho f + \lambda^r{}_{\beta\gamma}X_r f$ are linear combinations of $X_\beta f$. Hence $\lambda^i{}_{\beta\gamma} = M^i{}_r I^r{}_{\beta\gamma} = 0$ are the conditions for integrability. Because of the tensor character of $M^i{}_j$ and the circumstance that $|M^i{}_j| \neq 0$ we may write $I^i{}_{\beta\gamma} = 0$ in place of $\lambda^i{}_{\beta\gamma} = 0$ and for this reason we shall call $I^i{}_{\beta\gamma}$ an integrability tensor. When $I^i{}_{\beta\gamma} = 0$ there then exist $N - q$ independent solutions $f^i(x)$ which represent ∞^{N-q} subspaces $S^q(c^j)$ of q dimensions when equated to arbitrary constants c^i.

Let us now consider the more general non-holonomic case wherein $I^i{}_{\beta\gamma} \neq 0$ and therefore $X_\beta f = 0$ are not completely integrable. In this instance it becomes more convenient to discuss the Pfaffian system $dx^i - L^i{}_\rho\, dx^\rho = 0$ corresponding to $X_\beta f = 0$ rather than these latter equations themselves. These equations determine an elemental q-spread of directions dx^A at points of S^N and in the holonomic case the totality of these elemental q-spreads envelops the ∞^{N-q} subspaces $S^q(c^i)$. For the sake of generality this non-holonomic viewpoint will be retained henceforth and the holonomic theory will become the special case characterized by $I^i{}_{\beta\gamma} = 0$. Similar remarks apply to the Pfaffian system $dx^\alpha - L^\alpha{}_r\, dx^r = 0$ corresponding to the complementary equations $X_i f = 0$ whose integrability conditions are $I^\alpha{}_{jk} = 0$.

The question naturally arises as to what conditions may serve to assure the

existence of an admissible coordinate system y^A for which one set of the L's, say $L^a{}_j(y)$, vanishes. This is answered by the following

THEOREM 6.1: *The factor tensor conditions*

$$I^a{}_{jk} = X_j L^a{}_k - X_k L^a{}_j = 0 \tag{6.1}$$

$$L^a{}_r L^r{}_\beta = 0 \tag{6.2}$$

are necessary and sufficient for the existence of an admissible coordinate transformation from general coordinates x^A to normal coordinates y^A for which $L^a{}_j(y) = 0$.

PROOF: Since $I^a{}_{jk}$ and $L^a{}_r L^r{}_\beta$ are factor tensors, if there is to exist a coordinate system y for which $L^a{}_j(y) = 0$ it is certain that the conditions of the theorem must necessarily hold. To show their sufficiency we have in virtue of (6.1) that the system $X_j y = \partial_j y + L^\rho{}_j \partial_\rho y = 0$ admits q independent principal solutions $y^a(x)$ satisfying the initial conditions $y^a(x_\beta, x_0^k) = x^a$. Let $y^i(x)$ be any $N - q$ functions independent of y^a for which $| X_j y^i | \neq 0$. We assert that $y^A = y^A(x)$ constitutes a non-singular transformation to coordinates y for which $L^a{}_j(y) = 0$. From the transformation law $\bar{L}^a{}_r X_j \bar{x}^r = X_j \bar{x}^a$ it follows on identifying $\bar{x}^A$ with y^A that $L^a{}_j(y) = 0$. Furthermore the transformation $x \to y$ is admissible, for $| X_\beta y^a | = | \partial_\beta y^a + L^r{}_\beta \partial_r y |$ has the initial value $| \delta^a{}_\beta - L^a{}_r L^r{}_\beta |_{x^k = x_0^k} = | \delta^a{}_\beta |$ by (6.2). Finally from $| \partial_B y^A | = | X_\beta y^a | \cdot | X_j y^i |$ we see that $| \partial_B y^A | \neq 0$ so that the transformation $x \to y$ is non-singular. A similar proof holds for the complementary

THEOREM 6.2: *The conditions $I^i{}_{\beta\gamma} = X_\beta L^i{}_\gamma - X_\gamma L^i{}_\beta = 0$ and $L^i{}_\rho L^\rho{}_j = 0$ are necessary and sufficient for the existence of coordinates y^A for which $L^i{}_\beta(y) = 0$.*

7. AN AFFINE CONNECTION FOR S^N

By means of the equations (5.1) the factor components $l^A{}_{BC}$ of connection are uniquely determined by the components $L^A{}_{BC}$ and conversely these latter components are uniquely determined by a choice of $l^A{}_{BC}$. In this section we shall adopt this converse viewpoint and shall define the factor components $l^A{}_{BC}$ in the simplest way at our disposal. We first recall from section 5 that four sets of the connection components $l^A{}_{BC}$ are factor tensors. We now restrict these to be zero

$$(R_{7.1}) \qquad l^i{}_{\beta\gamma} = 0 \qquad l^a{}_{jk} = 0 \qquad l^i{}_{j\gamma} = 0 \qquad l^a{}_{\beta k} = 0.$$

From $\omega^A{}_{BC} = \tfrac{1}{2}[(l^A{}_{BC} - l^A{}_{CB}) - \lambda^A{}_{BC}]$ as given by (5.5) it follows as a result of $(R_{7.1})$ that the quantities $\omega^a{}_{j\gamma} = \tfrac{1}{2}(l^a{}_{j\gamma} - \lambda^a{}_{j\gamma})$ and $\omega^i{}_{\beta k} = \tfrac{1}{2}(l^i{}_{\beta k} - \lambda^i{}_{\beta k})$ are factor tensors. We therefore make the additional restrictions

$$(R_{7.2}) \qquad l^a{}_{j\gamma} = \lambda^a{}_{j\gamma} \qquad l^i{}_{\beta k} = \lambda^i{}_{\beta k}$$

which leave us with but two undefined sets of components $l^a{}_{\beta\gamma}$ and $l^i{}_{jk}$. Thus an entire affine connection for S^N is determined by $(R_{7.1})$, $(R_{7.2})$ and definitions of $l^a{}_{\beta\gamma}$ and $l^i{}_{jk}$.

The restrictions just imposed on the factor components of affine connection simplify the formulas of covariant differentiation as follows:

$$
\begin{aligned}
&\lambda_{\gamma,\beta} = X_\beta\lambda_\gamma - \lambda_\rho l^\rho{}_{\beta\gamma} && \lambda_{k,j} = X_j\lambda_k - \lambda_r l^r{}_{jk}\\
&\lambda_{\gamma,j} = X_j\lambda_\gamma - \lambda_\rho\lambda^\rho{}_{j\gamma} && \lambda_{k,\beta} = X_\beta\lambda_k - \lambda_r\lambda^r{}_{\beta k}\\
&\lambda^i{}_{,\beta} = X_\beta\lambda^i + \lambda^r\lambda^i{}_{\beta r} && \lambda^\alpha{}_{,j} = X_j\lambda^\alpha + \lambda^\rho\lambda^\alpha{}_{j\rho}\\
&\lambda^i{}_{,j} = X_j\lambda^i + \lambda^r l^i{}_{jr} && \lambda^\alpha{}_{,\beta} = X_\beta\lambda^\alpha + \lambda^\rho l^\alpha{}_{\beta\rho}.
\end{aligned}
\tag{7.3}
$$

We define parallel displacement of a contravariant vector Λ^A along a curve $x^A = x^A(s)$, where s is an affine parameter, by the equations $\Lambda^A{}_{,Q}\dfrac{dx^Q}{ds} = 0$. For a curve lying in a subspace S^q, holonomic or non-holonomic, the tangent vector $T^A = \dfrac{dx^A}{ds}$ with factor components $t^A = M^A{}_Q\dfrac{dx^Q}{ds}$ satisfies the Pfaffian system $t^i = M^i{}_r\left(\dfrac{dx^r}{ds} - L^i{}_\rho\dfrac{dx^\rho}{ds}\right) = 0$. From $\dfrac{dx^A}{ds} = L^A{}_Q t^Q$ it then follows that $t^\alpha = \dfrac{dx^\alpha}{ds}$. The factor components of $\Lambda^A{}_{,Q}\dfrac{dx^Q}{ds} = 0$ may be written $L^S{}_Q\partial_S\lambda^A M^Q{}_R\dfrac{dx^R}{ds} + l^A{}_{RS}t^R\lambda^S = 0$ or $\dfrac{d\lambda^A}{ds} + l^A{}_{RS}t^R\lambda^S = 0$ so that for parallel displacement along a curve in S^q we have

$$
\frac{d\lambda^\alpha}{ds} + l^\alpha{}_{\rho\sigma}\frac{dx^\rho}{ds}\lambda^\sigma = 0 \qquad \frac{d\lambda^i}{ds} + \lambda^i{}_{\rho r}\frac{dx^\rho}{ds}\lambda^r = 0 \qquad \frac{dx^i}{ds} - L^i{}_\rho\frac{dx^\rho}{ds} = 0. \tag{7.4}
$$

For parallel displacement of a covariant vector Λ_A in S^q the corresponding system $\Lambda_{A,Q}\dfrac{dx^Q}{ds} = 0$ has the factor components

$$
\frac{d\lambda_\alpha}{ds} - \lambda_\rho l^\rho{}_{\sigma\alpha}\frac{dx^\sigma}{ds} = 0 \qquad \frac{d\lambda^i}{ds} - \lambda_r\lambda^r{}_{\rho i}\frac{dx^\rho}{ds} = 0 \qquad \frac{dx^i}{ds} - L^i{}_\rho\frac{dx^\rho}{ds} = 0. \tag{7.5}
$$

A path of S^q is defined as a curve whose tangent vector $\lambda^\alpha = \dfrac{dx^\alpha}{ds}$, $\lambda^i = 0$, undergoes parallel displacement with respect to the curve. From (7.4) the equations of a path of S^q are

$$
\frac{d^2x^\alpha}{ds^2} + l^\alpha{}_{\rho\sigma}\frac{dx^\rho}{ds}\frac{dx^\sigma}{ds} = 0 \qquad \frac{dx^i}{ds} - L^i{}_\rho\frac{dx^\rho}{ds} = 0. \tag{7.6}
$$

8. The Riemannian S^N

In this section we shall consider only the completely holonomic case for which $\lambda^A{}_{\beta\gamma} = \lambda^A{}_{jk} = 0$ and shall assume that S^N is Riemannian with a metric tensor G_{AB}. Representing the Christoffel symbols by $\Gamma^A{}_{BC} = \frac{1}{2}G^{AQ}(\partial_B G_{CQ} + \partial_C G_{BQ} - \partial_Q G_{BC})$ and their factor components by $\left\{\begin{matrix} A \\ BC \end{matrix}\right\}$ these latter quantities will be given

in terms of $\Gamma^A{}_{BC}$ and $L^A{}_B$ by relations of the type (5.1). We desire the explicit form of $\left\{ {A \atop BC} \right\}$ in terms of the factor components $g_{AB} = G_{QR}L^Q{}_A L_{RB}$ of the metric tensor and for this purpose shall make use of the circumstance that $G_{AB,C} = 0$ and hence its frame components $g_{AB,C} = X_C g_{AB} - g_{QB}\left\{ {Q \atop CA} \right\} - g_{AQ}\left\{ {Q \atop CB} \right\}$ must likewise vanish. Substituting into $\frac{1}{2}(g_{AB,C} + g_{AC,B} - g_{BC,A}) = 0$ and using the relations $\omega^A{}_{BC} = \frac{1}{2}\left[\left(\left\{ {A \atop BC} \right\} - \left\{ {A \atop CB} \right\}\right) - \lambda^A{}_{BC}\right] = 0$ expressing the vanishing of the factor components of the skew-symmetric portion of $\Gamma^A{}_{BC}$ we obtain after minor rearrangements $\frac{1}{2}g_{AQ}\left(\left\{ {Q \atop BC} \right\} + \left\{ {Q \atop CB} \right\}\right) = [BC, A] - \frac{1}{2}(g_{BQ}\lambda^Q{}_{CA} + g_{CQ}\lambda^Q{}_{BA})$, where $[BC, A] = \frac{1}{2}(X_B g_{CA} + X_C g_{BA} - X_A g_{BC})$. Combining these with $\frac{1}{2}g_{AQ}\left(\left\{ {Q \atop BC} \right\} - \left\{ {Q \atop CB} \right\}\right) = \frac{1}{2}g_{AQ}\lambda^Q{}_{BC}$ we arrive at the desired expressions

$$g_{AQ}\left\{ {Q \atop BC} \right\} = [BC, A] - \tfrac{1}{2}(g_{BQ}\lambda^Q{}_{CA} + g_{CQ}\lambda^Q{}_{BA} - g_{AQ}\lambda^Q{}_{BC}). \tag{8.1}$$

We now ask if it is possible to restrict the metric tensor g_{AB} so that the quantities $\left\{ {A \atop BC} \right\}$ will satisfy the conditions $(R_{7.1})$ and $(R_{7.2})$ imposed in section 7 on the general affine connection $l^A{}_{BC}$. It may be verified that the conditions

$$(R_{8.2}) \qquad \lambda^i{}_{\beta\gamma} = 0 \qquad \lambda^\alpha{}_{jk} = 0 \qquad g_{i\alpha} = 0$$

in conjunction with $(R_{7.1})$ and $(R_{7.2})$ are sufficient to assure compatibility, for on substituting these values into (8.1) we are left with but two sets of non-vanishing relations $g_{\alpha\rho}\left\{ {\rho \atop \beta\gamma} \right\} = [\beta\gamma, \alpha]$ and $g_{ir}\left\{ {r \atop jk} \right\} = [jk, i]$ defining the components $\left\{ {\alpha \atop \beta\gamma} \right\}$ and $\left\{ {i \atop jk} \right\}$ which together with $(R_{7.1})$ and $(R_{7.2})$ define the complete factor components $\left\{ {A \atop BC} \right\}$ of the Christoffel symbols $\Gamma^A{}_{BC}$. Under these conditions there exist holonomic coordinates z^A for which $L^A{}_B(z) = \delta^A{}_B$ and the metric tensor $G_{AB}(z) = g_{AB}(z)$ is in the product form $\left\| \begin{matrix} G_{\alpha\beta}(z^\gamma) & 0 \\ 0 & G_{ij}(z^k) \end{matrix} \right\|$ as follows from $g_{\alpha\beta,k} = 0$ and $g_{ij,\gamma} = 0$ when written in normal form.

9. Contact Transformations

In the study of contact transformations a change in notation is desirable. Here we are concerned with $N = 2q$ variables x^A commonly denoted by x^i and p_i. We adapt the general factorization notation of the preceding sections to the contact group in $2q$ variables by choosing the two subranges to be $1, \cdots, q$ and $q + 1, \cdots, 2q$. Vector components will be written in the form (covariant)

$|| \Lambda_A || = ||\Lambda_i, \Lambda^i||$, (contravariant) $|| \Lambda^A || = || \Lambda^i, \Lambda_i ||$. This convention is in need of an additional notation to distinguish between covariant and contravariant quantities with indices in the same position but we shall deal only with a few tensors whose transformation character is clearly understood so that no ambiguity will arise. A similar scheme will be used for non-tensor quantities; for example we shall write $L^i{}_{q+j} = L^{ij}$ and $L^{q+i}{}_j = L_{ij}$.

The full contact transformation group in $2q$ variables x^i and p_i may be defined as the totality of transformations $X^i = X^i(x, p)$, $P_i = P_i(x, p)$ for which $P_r\,dX^r - p_r\,dx^r$ regarded as an expression in the $2q$ variables x^i and p_i and their differentials is the complete differential of some function of the x's and p's, that is,

$$P_r(x, p)\,dX^r(x, p) - p_r\,dx^r = dW(x, p). \tag{9.1}$$

More generally, a Pfaffian form $\varphi_Q(u)\,du^Q$ in $2q$ variables u^A will be said to be *preserved* modulo a complete differential under the transformation $u \to \bar{u}$ if the relation $(\varphi_Q(\bar{u})\partial_B\bar{u}^Q - \varphi_B(u))\,du^B = dW(u)$ holds, the necessary and sufficient conditions on the coefficients of du^B being $\varphi_{QR}(\bar{u})\partial_A\bar{u}^Q\partial_B\bar{u}^R = \varphi_{AB}(u)$, where $\varphi_{AB} = \partial_A\varphi_B - \partial_B\varphi_A$. From the general theory of Pfaffian forms the condition $|\varphi_{AB}| \neq 0$ is necessary and sufficient for the existence of a transformation $u \to x$ sending the Pfaffian $\varphi_Q\,du^Q$ into a canonical form $p_r\,dx^r$. A contact transformation $x^A \to \bar{x}^A$ may therefore be described as the most general transformation preserving the canonical form modulo a complete differential. For the canonical Pfaffian $\Pi_Q\,dx^Q = p_r\,dx^r$ the matrix of $\Pi_{AB} = \partial_A\Pi_B - \partial_B\Pi_A$ is simply

$$||\Pi_{AB}|| = \begin{Vmatrix} \Pi_{ij} & \Pi_i{}^j \\ \Pi^i{}_j & \Pi^{ij} \end{Vmatrix} = \begin{Vmatrix} 0 & -\delta_i^j \\ \delta^i_j & 0 \end{Vmatrix} \tag{9.2}$$

and the tensor transformation law $\Pi_{AQ}\bar{\partial}_B x^Q = \Pi_{QB}\partial_A\bar{x}^Q$ of its elements contains in concise form the familiar equations of definition of the contact group

$$-\bar{\partial}_j p_i = \partial_i\bar{p}_j \qquad -\bar{\partial}^j x^i = \partial^i\bar{x}^j \qquad \bar{\partial}^j p_i = \partial_i\bar{x}^j \tag{9.3}$$

as was first recognized by Eisenhart and Knebelman (1).

When the function $W(x, p)$ of (9.1) vanishes identically it may be shown that $P_i(x, p)$ and $X^i(x, p)$ are homogeneous in the p's of degrees one and zero respectively. For this reason transformations $x^A \to \bar{x}^A$ for which $\bar{p}_r\,d\bar{x}^r = p_r\,dx^r$ are called homogeneous contact transformations and they form a subgroup of the full contact group. By introducing two new variables, x^{q+1} and $p_{q+1} = -1$ and extending the transformation $x^A \to \bar{x}^A$ to

$$\bar{x}^i = \bar{x}^i(x, p) \qquad \bar{p}_i = \bar{p}_i(x^j, p_k)$$
$$\bar{x}^{q+1} = x^{q+1} + W(x^i, p_j) \qquad \bar{p}_{q+1} = p_{q+1} = -1$$

(9.1) becomes $\sum_1^{q+1} \bar{p}_\rho\,d\bar{x}^\rho = \sum_1^{q+1} p_\rho\,dx^\rho$. Since any contact transformation in $2q$ variables may be written in this way as a homogeneous contact transformation in $2q + 2$ variables we shall restrict our discussion henceforth to homogeneous transformations in $2q$ variables.

The coefficients $\| \Pi_A \| = \| p_i, 0 \|$ of the Pfaffian $p_r\, dx^r$ form a vector whose transformation law $\Pi_Q \partial_A \bar{x}^Q = \Pi_A$ gives the relations

$$P_r \partial_i X^r = p_i \qquad P_r \partial^i X^r = 0 \tag{9.4}$$

which are equivalent to (9.3) plus the homogeneity relations $p_r \partial^r X^i = 0$ and $p_r \partial^r P_i = P_i$ and either set fully characterizes the homogeneous contact group. We shall need the factor components $\pi_A = \Pi_Q L^Q{}_A$ and $\pi_{AB} = \Pi_{QR} L^Q{}_A L^R{}_B$ for future reference. They are

$$\| \pi_A \| = \| p_i, p_r L^{ri} \|$$

$$\| \pi_{AB} \| = \begin{Vmatrix} \pi_{ij} & \pi_i{}^j \\ \pi^i{}_j & \pi^{ij} \end{Vmatrix} = \begin{Vmatrix} -(L_{ij} - L_{ji}) & -(\delta^j{}_i - L_{ri} L^{rj}) \\ (\delta^i{}_j - L_{rj} L^{ri}) & (L^{ij} - L^{ji}) \end{Vmatrix}. \tag{9.5}$$

10. Normal Coordinates for the Contact Frame

Theorem 10.1: *The factor tensor conditions*

$$I^{ijk} = X^j L^{ik} - X^k L^{ij} = 0 \quad \pi^{ij} = L^{ij} - L^{ji} = 0 \quad \pi^i = p_r L^{ri} = 0 \tag{R$_{10.1}$}$$

are necessary and sufficient for the existence of a homogeneous contact transformation from general coordinates x^A to normal coordinates y^A for which $L^{ij}(y) = 0$.

Proof: The necessity of these conditions is obvious from their linearity in L^{ij} and their tensor character. To prove their sufficiency we notice from the transformation equations $\bar{L}^{jr} X^k \bar{p}_r = X^k \bar{x}^j$ of the frame components L^{jk} that if we identify $\bar{x}^A$ with the desired normal coordinates y^A we must choose $Y^i(x, p)$ as solutions of the system $X^k Y = \partial^k Y + L^{rk} \partial_r Y = 0$. The conditions $I^{ijk} = 0$ assure complete integrability of these equations so that we may find q functionally independent solutions $Y^i(x, p)$ satisfying

$$\partial^k Y^i + L^{rk} \partial_r Y^i = 0, \tag{10.2}$$

and by the theory of section 1 for equations of this solved type the solutions Y^i satisfy $| \partial_k Y^i | \neq 0$. Contraction of equations (10.2) with p_k shows that the solutions Y^i are homogeneous in the p's of degree zero. Because $| \partial_j Y^i | \neq 0$ we may define q functions $Q_i(x, p)$ as the unique solutions of $Q_r \partial_j Y^r = p_j$ which will be homogeneous in the p's of degree one. By the contraction $Q_r(\partial^k Y^r + L^{sk} \partial_s Y^r) = 0$ we find $Q_r \partial^k Y^r = 0$ and hence $y^i = Y^i(x, p)$ and $q^i = Q_i(x, p)$ satisfy the conditions (9.4) for a homogeneous contact transformation.

We next ask for the most general contact transformation preserving normal coordinates. By identifying x^A with y^A and $\bar{x}^A$ with $\bar{y}^A$ in $\bar{L}^{ir} X^j \bar{p}_r = L^{rj} X_r \bar{x}^i + t_r{}^{ij} \partial^r \bar{x}^i$ as given by (4.4), it follows immediately that any transformation from normal coordinates y for which $L^{ij}(y) = 0$ to a system $\bar{y}$ for which also $\bar{L}^{ij}(\bar{y}) = 0$ must satisfy $\dfrac{\partial \bar{y}^i}{\partial q_j} = 0$ and therefore must be an extended point transformation. Hence the

Theorem 10.2: *The most general homogeneous contact transformation $y^A \to \bar{y}^A$ preserving normal coordinates for which $L^{ij} = 0$ is an extended point transformation.*

11. Restrictions on the Contact Frame

We shall impose henceforth the restrictions ($R_{10.1}$) assuring the existence of normal coordinates. If we define a *union* to be any subspace of S^{2q} within which an arbitrary displacement dx^A at a point (x^A) satisfies $p_r\,dx^r = 0$ it may be shown that such a subspace may be regarded as a point locus in the point space of coordinates x^i together with all its tangent hyperplane elements with coordinates p_i. Thus the maximum dimensionality of a union is $q - 1$. In terms of unions the restrictions ($R_{10.1}$) require that any q independent solutions of the system $X^i f = 0$ when equated to arbitrary constants shall constitute a union of maximum dimensionality $q - 1$, for in normal coordinates the system becomes $\dfrac{\partial f}{\partial q_i} = 0$ and its solutions y^i determine the ∞^q unions $y^i = y^i_0$ consisting of the point (y^i_0) in the point space Y^q regarded as the envelope of its ∞^{q-1} hyperplane elements q_i.

By operating on the transformation laws (2.2) with the invariant operator $p_r\partial^r$ it may be shown that the quantities $h_{jk} = L_{jk} - p_r\partial^r L_{jk}$ and $h^{jk} = L^{jk} + p_r\partial^r L^{jk}$ are covariant factor tensors, and since h^{jk} vanishes in normal coordinates, $L^{jk} + p_r\partial^r L^{jk} = 0$. As a second set of restrictions on the contact frame we introduce

$$(R_{11.1}) \qquad -\pi_{jk} = L_{jk} - L_{kj} = 0 \qquad h_{jk} = L_{jk} - p_r\partial^r L_{jk} = 0.$$

We next collect for future reference some of the more useful consequences of the restrictions of this section. From equations (2.3) we have $\bar{L}_{rk}\bar{\partial}^r x^i = -L^{ir}\bar{\partial}_j p_r$ and hence $X_j\bar{x}^i = \bar{X}^i p_j$. The transformation laws of factor vectors simplify to $\bar{\lambda}_j = \lambda_r \bar{X}_j x^r$ and $\bar{\lambda}^j = \lambda^r X_r \bar{x}^j$ for both the covariant vector $\| \lambda_A \| = \| \lambda_i , \lambda^i \|$ and the contravariant vector $\| \lambda^A \| = \| \lambda^i, \lambda_i \|$ so that an upper index of covariance has contravariant character and a lower index of contravariance has covariant character. Under extended point transformations $\bar{y}^i = \bar{y}^i(y^j)$, $\bar{q}_i = q_r\bar{\partial}_i y^r$, these laws reduce to the familiar form $\bar{\lambda}_j(\bar{y}) = \lambda_r(y)\bar{\partial}_j y^r$, $\lambda^j(\bar{y}) = \lambda^r(y)\partial_r \bar{y}^j$. The two tensors $L^i{}_\rho L^\rho{}_j$ and $L^\alpha{}_r L^r{}_\beta$ of section 4 coincide to give just one vanishing tensor $L^{ir}L_{rj} = 0$ and the two mixed tensors $t^\alpha{}_\beta$ and $t^i{}_j$ reduce to the single Kroencker delta $t^i{}_j = \delta^i{}_j$. The transformation laws (4.4) of the L's are now

$$(11.2) \qquad \bar{L}_{ir}\partial_j \bar{x}^r = L_{rj}\bar{\partial}_i x^r + \partial_j \bar{p}_i \qquad \bar{L}^{ir}\bar{\partial}_r x^j = L^{rj}\partial_r \bar{x}^i + \partial^j \bar{x}^i$$

and from (9.5) the factor tensors π_A and π_{AB} have the same components as Π_A and Π_{AB}, $\| \pi_A \| = \| p_i , 0 \|$, $\| \pi_{AB} \| = \begin{Vmatrix} 0 & -\delta_i{}^j \\ \delta^i{}_j & 0 \end{Vmatrix}$. The matrix $\| M^A{}_B \|$ conjugate to $\| L^A{}_B \|$ is $\begin{Vmatrix} M^i{}_j & M^{ij} \\ M_{ij} & M_i{}^j \end{Vmatrix} = \begin{Vmatrix} \delta^i_j & -L^{ij} \\ -L_{ij} & \delta_i^j \end{Vmatrix}$. In normal coordinates y^A the commutator symbols $\lambda^A{}_{BC}$ have the values $\lambda^i{}_{jk}(y) = \lambda^{ijk}(y) = \lambda^{ij}{}_k(y) = 0$, $\lambda_{ijk}(y) = Y_j L_{ik} - Y_k L_{ij}$, $\lambda_{ij}{}^k(y) = -\partial^k L_{ij}$.

12. A Complete Determination of an Affine Connection for Contact Transformations

We carry over into this section the restrictions ($R_{7.1}$) and ($R_{7.2}$) of the general theory of section 7. In our present notation these restrictions are

$$(R_{12.1}) \qquad l_{ijk} = l^{ijk} = l_i{}^j{}_k = l^i{}_j{}^k = 0, \qquad l^{ij}{}_k = \lambda^{ij}{}_k, \qquad l_{ij}{}^k = \lambda_{ij}{}^k,$$

and so a complete determination of factor components of affine connection will result when we have selected the two sets of components $l^i{}_{jk}$ and $l_i{}^{jk}$. This choice will be made by insisting that the covariant derivative of the factor tensor π_{AB} be zero. From (7.3) the only conditions are $X_j\pi_k{}^i - \pi_r{}^i l^r{}_{jk} - \pi_k{}^r\lambda_{rj}{}^i = 0$ and $X^j\pi_i{}^k - \pi_r{}^k\lambda^{rj}{}_i - \pi_k{}^r l_r{}^{jk} = 0$, which reduce by the values of the π's given in section 11 to

$$(R_{12.2}) \qquad l^i{}_{jk} + \lambda_{kj}{}^i = 0 \qquad l_i{}^{jk} + \lambda^{kj}{}_i = 0.$$

These restrictions define the components $l^i{}_{jk}$ and $l_i{}^{jk}$ so that a complete affine connection is given by ($R_{12.1}$) and ($R_{12.2}$).

On expanding $\lambda_{kj}{}^i$ and $\lambda^{kj}{}_i$ as defined in section 5 we find $\lambda_{kj}{}^i = -\partial^i L_{kj} + L^{ir}I_{kjr}$ and $\lambda^{kj}{}_i = -\partial_i L^{kj} + L_{ir}I^{kjr} = -\partial_i L^{kj}$ so that $l^i{}_{jk} = \partial^i L_{kj} - L^{ir}I_{kjr}$ and $l_i{}^{jk} = \partial_i L^{jk}$. By contraction with p_i we find $p_r l^r{}_{jk} = L_{jk}$ which will be interpreted in section (14). The formulas of covariant differentiation coincide in pairs to give

$$\lambda_{k,j} = X_j\lambda_k - \lambda_r l^r{}_{jk} \qquad \lambda^{k,j} = X^j\lambda^k + \lambda^r\lambda^{kj}{}_r$$

$$\lambda_k{}^{,j} = X^j\lambda_k - \lambda_r\lambda^{rj}{}_k \qquad \lambda^k{}_{,j} = X_j\lambda^k + \lambda^r l^k{}_{jr},$$

where the vector λ is either covariant with components $(\lambda_A) = (\lambda_i, \lambda^i)$ or contravariant with components $(\lambda^A) = (\lambda^i, \lambda_i)$.

13. Holonomic Coordinates for the Contact Frame

If we operate on (5.1), $X_D L^A{}_B = l^Q{}_{DB}L^A{}_Q - L^A{}_{QR}L^Q{}_D L^R{}_B$ with $X_C = L^Q{}_C\partial_Q$ and then form $(X_C X_D)L^A{}_B = \lambda^Q{}_{CD}X_Q L^A{}_B$ we obtain $l^Q{}_{BCD}L^A{}_Q = L^A{}_{QRS}L^Q{}_B L^R{}_C L^S{}_D$, where $L^A{}_{BCD}$ is the curvature tensor $L^A{}_{BCD} = \partial_C L^A{}_{DB} - \partial_D L^A{}_{CB} + L^A{}_{CQ}L^Q{}_{DB} - L^A{}_{DQ}L^Q{}_{CB}$ and the quantities $l^A{}_{BCD} = X_C l^A{}_{DB} - X_D l^A{}_{CB} + l^A{}_{CQ}l^Q{}_{DB} - l^A{}_{DQ}l^Q{}_{CB} - \lambda^Q{}_{CD}l^A{}_{QB}$ are its factor components. Under the restrictions of sections 11 and 12 the only non-vanishing factor components of the contact curvature tensor are

$$l^i{}_{jkl} = -l_j{}^i{}_{kl} = X_l\lambda_{jk}{}^i - X_k\lambda_{jl}{}^i + \lambda_{rk}{}^i\lambda_{jl}{}^r - \lambda_{rl}{}^i\lambda_{jk}{}^r + \lambda^r{}_{kl}\lambda_{jr}{}^i - \lambda_{rkl}\lambda^{ir}{}_j$$

$$l^i{}_{jk}{}^l = \quad l_j{}^{il}{}_k = X_k\lambda^{il}{}_j + X^l\lambda_{jk}{}^i + \lambda^{il}{}_r\lambda_{jk}{}^r - \lambda_{rk}{}^i\lambda^{rl}{}_j - \lambda^{rl}{}_k\lambda_{jr}{}^i - \lambda_{rk}{}^l\lambda^{ir}{}_j$$

and they reduce in normal coordinates y^A to

$$l^i{}_{jkl}(y) = \partial_k(\partial^i L_{lj}) - \partial_l(\partial^i L_{kj}) + \partial^i L_{kr}\partial^r L_{lj} - \partial^i L_{lr}\partial^r L_{kj}$$

(13.1) $$+ L_{rk}\partial^{2ri}L_{lj} - L_{rl}\partial^{2ri}L_{kj}$$

$$l^{i}{}_{jk}{}^{l}(y) = -\partial^{2li}L_{jk}\,.$$

We next seek the conditions for the existence of holonomic coordinates $(z^A) = (z^i, r_i)$ for which simultaneously $L^{jk}(z) = 0$ and $L_{jk}(z) = 0$.

THEOREM 13.1: *The conditions*

(13.2) $$I^{ijk} = X^jL^{ik} - X^kL^{ij} = 0 \qquad \pi^{ij} = L^{ij} - L^{ji} = 0 \quad \pi^i = p_rL^{ri} = 0$$

(13.3) $$I_{ijk} = X_jL_{ik} - X_kL_{ij} = 0 \quad -\pi_{ij} = L_{ij} - L_{ji} = 0 \quad l^{i}{}_{jk}{}^{l} = 0$$

are necessary and sufficient for the existence of a homogeneous contact transformation from general coordinates x^A to holonomic coordinates z^A for which $L^A{}_B(z) = \delta^A{}_B$.

PROOF: From theorem 10.1 the first set of conditions is necessary and sufficient for the existence of normal coordinates y^A for which $L^{ij}(y) = 0$. In this coordinate system the left members of the last of (13.3) reduce to (13.1) so that if there is to exist a coordinate system z for which $L_{jk}(z) = 0$ it is at once obvious that the conditions (13.3) must necessarily hold. The sufficiency of the conditions of the theorem may be established by first applying theorem 10.1 which asserts the sufficiency of conditions (13.1) for the existence of normal coordinates y^A. We then apply theorem (10.2) which states that the most general homogeneous contact transformation $y \to z$ preserving $L^{ij} = 0$ is an extended point transformation,

$$T\colon\ z^i = z^i(y^j) \qquad r_i = q_r\frac{\partial y^r}{\partial z^i} \qquad T^{-1}\colon\ y^i = y^i(z^j) \qquad q_i = r_s\frac{\partial z^s}{\partial y^i}.$$

It remains to be shown that the conditions (13.3) are sufficient for the existence of q independent functions $z^i(y^j)$ such that T will constitute a transformation to coordinates z^A for which $L_{jk}(z) = 0$ as well as $L^{jk}(z) = 0$. To accomplish this we note that the transformation law $L_{jr}(y)\left(\frac{\partial y^r}{\partial z^k} + L_{sk}(z)\frac{\partial y^r}{\partial r_s}\right) = \frac{\partial q_j}{\partial z^k} + L_{sk}(z)\frac{\partial q_j}{\partial r_s}$ reduces to $L_{jr}(y)\frac{\partial y^r}{\partial z^k} = q_i\frac{\partial y^i}{\partial z^s}\frac{\partial}{\partial y^r}\left(\frac{\partial z^s}{\partial y^j}\right)\frac{\partial y^r}{\partial z^k}$ for an extended point transformation T, and since $\left|\frac{\partial y^r}{\partial z^k}\right| \neq 0$, $L_{jk}(y) = q_i\frac{\partial y^i}{\partial z^s}\frac{\partial}{\partial y^k}\left(\frac{\partial z^s}{\partial y^j}\right)$. Differentiating with respect to q_i and solving for $\frac{\partial}{\partial y^k}\left(\frac{\partial z^i}{\partial y^j}\right)$ we obtain $\frac{\partial}{y^k}\left(\frac{\partial z^i}{\partial y^j}\right) = \frac{\partial z^i}{\partial y^r}\partial^r L_{jk}$ and thus arrive at the linear system

(13.4) $$\partial_j z^i = Z_j{}^i(y^k)$$

(13.5) $$\partial_k Z_j{}^i = Z_r{}^i\partial^r L_{jk}$$

in $q(q+1)$ unknowns z^i and $Z_j{}^i$ to be determined as functions of the q variables y^i. Since the quantities $Z_j{}^i$ may be functions of only y^i it is necessary that $\partial^{2il}L_{jk}(y) = 0$ and these are satisfied by the last of (13.3). The integrability

conditions of (13.4) are found by expressing the conditions that $Z_r{}^i(\partial^r L_{jk} - \partial^r L_{kj})$ shall vanish for all $Z_j{}^i$, namely $\partial^i(L_{jk} - L_{kj}) = 0$, and these are satisfied by the middle set of (13.3). Finally the integrability conditions of (13.5) are $\partial_l(\partial^i L_{jk}) - \partial_k(\partial^i L_{jl}) + \partial^i L_{rl}\partial^r L_{jk} - \partial^i L_{rk}\partial^r L_{jk} = 0$ and these are satisfied by the first of (13.3) as may be seen by differentiating in normal coordinates with respect to q_i. Finally from the general theory of completely integrable systems of the type (13.4) and (13.5) there exist solutions $Z_j{}^i$ for which $|Z_j{}^i| = |\partial_j z^i| \neq 0$. This completes the proof of the theorem.

14. Introduction of a Metric Tensor

The trajectories of a one parameter homogeneous contact group are solutions $x^A(s)$ of the system

$$\frac{dx^i}{ds} = \partial^i C \qquad \frac{\partial p_i}{ds} = -\partial_i C \tag{14.1}$$

whose vector character is obvious when written in the form $\Pi_{AQ}\dfrac{dx^Q}{ds} = \partial_A C$. $C(x, p)$ is necessarily homogeneous of degree one in the p's and hence we may take $C = 1$ providing we exclude the singular locus $C = 0$ from our discussion. Since C is an integral of (14.1) the condition $C = 1$ will hold at all points of a trajectory if it holds at one point.

We now choose the scalar C to be a solution of $X_j f = 0$,

$$\partial_j C + L_{rj}\partial^r C = 0. \tag{R$_{14.2}$}$$

Then the second set of (14.1) becomes $\dfrac{dp_i}{ds} = L_{ri}\dfrac{dx^r}{ds}$. These equations have been discussed in section 7 where it was seen that they express the vanishing of λ_i as defined by $\dfrac{dx^A}{ds} = L^A{}_Q\lambda^Q$ so that these latter reduce to $\lambda^i = \dfrac{dx^i}{ds}$ and $\lambda_i = \dfrac{dp_i}{ds} - L_{ir}\dfrac{dx^r}{ds} = 0$. Thus the displacement dx^A is restricted to the non-holonomic subspace X^q of S^{2q}, where X^q is the point space of points (x^i) and is characterized as a non-holonomic subspace of S^{2q} by the arbitrariness of the differentials dx^i. Furthermore it may be seen from (7.5) that the equations of parallel displacement of the vector $(\pi_A) = (p_i, 0)$ are $\dfrac{dp_i}{ds} - p_r\Gamma^r{}_{si}\dfrac{dx^s}{ds} = 0$, but from the last of section 12 these become $\lambda_i = 0$ so that the trajectories under consideration have the property that the vector π_A undergoes parallel displacement along them.

We shall now introduce a metric tensor $G_{AB}(x)$, whose factor components g_{AB} will be of the restricted form $\|g_{AB}\| = \begin{Vmatrix} g_{ij} & 0 \\ 0 & g^{ij} \end{Vmatrix}$ where $g_{ir}g^{rj} = \delta_i{}^j$. We define g^{ij} in terms of the scalar $H = \frac{1}{2}C^2$, where C is the scalar satisfying $X_j C = 0$, by $g^{ij}(x, p) = (X^i H)^{\cdot k}$ reducing in normal coordinates to $g^{ij}(y, q) = \partial^{2ij}H(y, q)$. From these definitions the components $g^{ij}(x, p)$ are homogeneous

of degree zero in the p's. There is a theorem (4) which states that a necessary and sufficient condition for the Hessian determinant $|\partial^{2ij}H|$ to be non-singular is that $|\partial^{2ij}C|$ be of maximum rank $q-1$. Restricting our discussion to those solutions C of $X_jf=0$ for which the rank of $|\partial^{2ij}C|$ is $q-1$ we have the two fundamental requirements, $g^{ij}=g^{ji}$ and $|g^{ij}|\neq 0$, of a metric tensor fulfilled. On doubly contracting the tensor g^{ij} with the covariant factor vector $\pi_i=p_i$ we obtain the invariant $g^{rs}p_rp_s$ which in normal coordinates is $\partial^{2rs}Hq_rq_s=2H$ and hence in all coordinates $H=\frac{1}{2}g^{rs}p_rp_s$. The system of trajectories (14.1) which may equally well be written $\frac{dx^i}{ds}=\partial^iH$ and $\frac{dp_i}{ds}=-\partial_iH$ is now

$$\frac{dx^i}{ds}=g^{ir}p_r \qquad \frac{dp_i}{ds}=-\tfrac{1}{2}\partial_ig^{rs}p_rp_s, \tag{14.3}$$

or in alternative form

$$p_i=g_{ir}\frac{dx^r}{ds} \qquad \frac{dp_i}{ds}=L_{ir}\frac{dx^r}{ds}. \tag{14.4}$$

We next seek additional restrictions on the frame $L^A{}_B$ and the metric g^{ij} which will identify the trajectories (14.3) with the geodesics of the subspace X^q. From equations (7.6) defining the paths of the point space X^q we arrive at the equations of the geodesics by substituting the factor components of the Christoffel symbols in place of the affine connections components. Thus the geodesics are given by

$$\frac{d^2x^i}{ds^2}+\left\{\begin{matrix} i \\ rs \end{matrix}\right\}\frac{dx^r}{ds}\frac{dx^s}{ds}=0 \qquad \frac{dp_i}{ds}-L_{ir}\frac{dx^r}{ds}=0 \tag{14.5}$$

and it is this system that is to be identified with (14.3).

15. Additional Restrictions on the Contact Frame

We now investigate the possibility of restricting g^{ij} to be free of the q's in normal coordinates, $\partial^kg^{ij}(y^A)=0$, as expressed by $g^{ij,k}=0$. This will necessitate that $H(y^A)=\frac{1}{2}g^{rs}(y^i)q_rq_s$ be a solution of $\partial_iH+L_{ri}\partial^rH=0$. The conditions are

$$\tfrac{1}{2}\partial_ig^{rs}q_rq_s=-L_{ir}g^{rs}q_s \tag{15.1}$$

and by repeated differentiation we derive the necessary conditions

$$\partial_ig^{jk}=-(\partial^kL_{ir}g^{rj}+\partial^jL_{ir}g^{rk})-\partial^{2jk}L_{ir}g^{rs}q_s.$$

If we introduce the curvature tensor restrictions

$$l^i{}_{jk}{}^l=0 \tag{$R_{15.2}$}$$

reducing in normal coordinates to $\partial^{2il}L_{jk}(y)=0$ we are left with

$$\partial_ig^{jk}=-(\partial^kL_{ir}g^{rj}+\partial^jL_{ir}g^{rk}) \tag{15.3}$$

which may be solved for the quantities $\partial^i L_{jk}$ as Christoffel symbols formed from $G_{ij}(y^A) = g_{ij}(y^k)$, $\partial^i L_{jk}(y) = \Gamma^i{}_{jk}(y^l)$, and hence

$$L_{ij}(y^l, q_m) = q_r \Gamma^r{}_{jk}(y^l). \tag{15.4}$$

Conversely, if the expressions (15.3) resulting from (15.4) by differentiation are substituted into (15.1) it may be verified that these latter equations are satisfied.

The restrictions just imposed on g^{ij} and L_{jk} are sufficient to bring the geodesics (14.5) into coincidence with the trajectories (14.3), for differentiation of the first of (14.3) in normal coordinates gives

$$\frac{d^2 y^i}{ds^2} - \Gamma^i{}_{rs}(y^j) \frac{dy^r}{ds} \frac{dy^s}{ds} = 0 \qquad \frac{dq_i}{ds} = L_{ir}(y^A) \frac{dy^r}{ds}. \tag{15.5}$$

Under the present restrictions the equations (8.1) define $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$ by $\left\{ \begin{matrix} i \\ jk \end{matrix} \right\} = \frac{1}{2} g^{ir}[(X_j g_{kr} + X_k g_{jr} - X_r g_{jk}) - (g_{js}\lambda^s{}_{kr} + g_{ks}\lambda^s{}_{jr} - g_{js}\lambda^s{}_{jk})]$ which reduce in normal coordinates to the familiar Christoffel symbols, $\{{}^i{}_{jk}(y)\} = \Gamma^i{}_{jk}(y^l)$. Thus the normal form of the equations (14.5) of the geodesics is likewise given by (15.5) and because of the vector character of (14.3) and (14.5) this completes the proof of their equivalence. It is interesting to note that from the first of (14.4) and the relations $p_r \lambda^r{}_{jk} = 0$ resulting from the definitions $\lambda^i{}_{jk} = -L^{ir} I_{rjk}$ of section 5, the first of (14.5) is given by $\dfrac{d^2 x^i}{ds} + g^{it}[\frac{1}{2}(X_r g_{st} + X_s g_{rt} - X_t g_{rs})] \dfrac{dx^r}{ds} \dfrac{dx^s}{ds} = 0$. The tensor $l^i{}_{jk} - \left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$ reduces in normal coordinates to $\partial^i L_{jk} - \Gamma^i{}_{jk}(y^l) = 0$ and hence in all coordinates $l^i{}_{jk} = \left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$. The paths of X^q are therefore identical with the geodesics. We shall regard the affine connection of S^{2q} to be that of section 12 where the components $l^i{}_{jk}$ as there defined now satisfy the additional relations $l^i{}_{jk} = \left\{ \begin{matrix} i \\ jk \end{matrix} \right\}$. Because of ($R_{15.2}$) the curvature tensor $l^A{}_{BCD}$ has but one set of non-vanishing components $l^i{}_{jkl} = -l_j{}^i{}_{kl}$ reducing in normal coordinates to the Riemann curvature tensor $R^i{}_{jkl}(y^m) = \partial_k \Gamma^i{}_{lj} - \partial_l \Gamma^i{}_{kj} + \Gamma^i{}_{kr}\Gamma^r{}_{lj} - \Gamma^i{}_{lr}\Gamma^r{}_{kj}$ formed from the Christoffel symbols $\Gamma^i{}_{jk}(y^l)$.

The results of this section may be summarized by the statement that the three ostensibly distinct systems of invariantive curves consisting of

1) the trajectories of $\Pi_{AQ} \dfrac{dx^Q}{ds} = \partial_A C$ where $X_j C = 0$,

2) the geodesics of the point space X^q with respect to the metric tensor $g^{ij}(x, p) = \frac{1}{2}(X^i C^2)^{\cdot j}$,

3) the paths of X^q relative to the connection $l^i{}_{jk} = \partial^i L_{kj} - L^{ir} I_{kjr}$ have been combined into just one system.

16. Doubly Homogeneous Contact Transformations

Let Σ be a projective space of $N - 1$ dimensions, S an arbitrary hypersurface of Σ parameterized by the equations $x^i = \varphi^i(u^\alpha)$, $i = 1, 2, \cdots, N$, $\alpha =$

$1, 2, \cdots, N-2$, and P an arbitrary point of S. The conditions on a set of hyperplane coordinates p_i that they represent the tangent hyperplane at P may be found by insisting that the plane p_i cut an arbitrary curve $u^\alpha = u^\alpha(t)$ in two coincident points at P whose parameter we choose to be $t = 0$. This means that $p_r x^r(t) = at^2 + \cdots$ when written as a power series in t. Hence the conditions for $t = 0$ are $p_r x^r = 0$ and $p_r \dfrac{\partial x^r}{\partial u^\alpha}\dfrac{du^\alpha}{dt} = 0$. But since these must hold for an arbitrary curve $u^\alpha(t)$ the conditions on the point-hyperplane element $(x^i(u), p_i(u))$ that it represent a point $x^i(u)$ incident with the plane $p_i(u)$ which in turn is tangent to the hypersurface S at x^i are $p_r x^r = 0$ and $p_r \dfrac{\partial x^r}{\partial u^\alpha} = 0$. But dually we may regard S as a locus of hyperplanes $p_i(u)$ and ask for the conditions on a point x^i that it be a "tangent point" of S. By analogous reasoning these conditions are $p_r x^r = 0$ and $x^r \dfrac{\partial p_r}{\partial u^\alpha} = 0$. An $N-2$ spread of point-hyperplane elements $(x^i(u), p_i(u))$ will be said to constitute a *union* after the terminology of Lie if the conditions $p_r x^r = 0$, $p_r \dfrac{\partial x^r}{\partial u^\alpha} = 0$, $x^r \dfrac{\partial p_r}{\partial u^\alpha} = 0$ are satisfied as identities in the arbitrary parameters u^α.

Suppose now that the union of elements $x^A(u)$ viewed in Σ either as a point locus $x^i = x^i(u)$ or as a hyperplane locus $p_i(u)$ is subjected to a transformation $\bar{x}^A = \bar{x}^A(x^B)$. If the new elements as functions of the same parameters u^α are to form a union with the point locus $\bar{x}^i = \bar{x}^i(x^B(u))$ and hyperplane locus $\bar{p}_i = \bar{p}_i(x^B(u))$ as a consequence of being the transform of a union we must have $\bar{p}_r \bar{x}_r = \lambda(x, p) p_r x^r$, $\bar{p}_r \dfrac{\partial \bar{x}^r}{\partial u^\alpha} = \mu(x, p) p_r \dfrac{\partial x^r}{\partial u^\alpha}$, $\bar{x}^r \dfrac{\partial \bar{p}}{\partial u^\alpha} = \nu(x, p) x^r \dfrac{\partial p_r}{\partial u^\alpha}$ for an arbitrary parameterization u^α. Expanding the last two conditions we have $\bar{p}_r \partial_i \bar{x}^r = p_i$, $\bar{p}_r \partial^i \bar{x}^r = 0$, $\bar{x}^r \partial^i \bar{p}_r = \nu x^i$, $\bar{x}^r \partial_i \bar{p}_r = 0$, and differentiation of the first with respect to x^i and p_i gives $(\lambda - \mu) p_i + \partial_i \lambda p_r x^r = 0$ and $(\lambda - \nu) x^i + \partial^i \lambda p_r x^r = 0$. But these latter conditions must hold identically both when $p_r x^r = 0$ and when $p_r x^r \neq 0$ so that $\partial_A \lambda = 0$ and $\lambda = \mu = \nu =$ constant. There will be two distinct types of such transformations depending on whether $\lambda > 0$ or $\lambda < 0$. In all generality we may take $\lambda = 1$ in the first case and $\lambda = -1$ in the second giving the two types

$$(\mathrm{I}_a) \quad \text{(Proper contact transformation)} \qquad \begin{matrix} \bar{p}_r \partial_i \bar{x}^r = p_i \\ \bar{p}_r \partial^i \bar{x}^r = 0 \end{matrix} \qquad \bar{p}_r \bar{x}^r = p_r x^r \qquad \begin{matrix} \bar{x}^r \partial^i \bar{p}_r = x^i \\ \bar{x}^r \partial_i \bar{p}_r = 0 \end{matrix}$$

$$(\mathrm{II}_a) \quad \text{(Improper contact transformation)} \qquad \begin{matrix} \bar{p}_r \partial_i \bar{x}^r = -p_i \\ \bar{p}_r \partial^i \bar{x}^r = 0 \end{matrix} \qquad \bar{p}_r \bar{x}^r = -p_r x^r \qquad \begin{matrix} \bar{x}^r \partial^i \bar{p}_r = -x^i \\ \bar{x}^r \partial_i \bar{p}_r = 0. \end{matrix}$$

Let us next consider an element transformation sending the point locus $x^i = x^i(u)$ into the hyperplane locus $\bar{p}_i = \bar{p}_i(x^B(u))$, and dually, the same locus when regarded as a hyperplane locus $p_i = p_i(u)$ into the same barred locus

regarded as a point locus $\bar{x}^i = \bar{x}^i(x^B(u))$. The conditions that the transformed elements $\bar{x}^A(u)$ constitute a union if the original elements $x^A(u)$ form a union are $\bar{p}_r\bar{x}^r = \lambda' p_r x^r$, $\bar{p}_r \dfrac{\partial \bar{x}^r}{\partial u^a} = \mu' x^r \dfrac{\partial p_r}{\partial u^a}$, $\bar{x}^r \dfrac{\partial \bar{p}_r}{\partial u^a} = \nu' p_r \dfrac{\partial x^r}{\partial u^a}$, and again it follows that $\lambda' = \mu' = \nu' =$ constant so that the transformations under discussion fall into two classifications:

(I_b) (Proper contact correlation)
$$\bar{p}_r \partial_i \bar{x}^r = 0 \qquad \bar{p}_r \bar{x}^r = p_r x^r \qquad \bar{x}^r \partial^i \bar{p}_r = 0$$
$$\bar{p}_r \partial^i \bar{x}^r = x^i \qquad\qquad \bar{x}^r \partial_i \bar{p}_r = p_i$$

(II_b) (Improper contact correlation)
$$\bar{p}_r \partial_i \bar{x}^r = 0 \qquad \bar{p}_r \bar{x}^r = -p_r x^r \qquad \bar{x}^r \partial^i \bar{p}_r = 0$$
$$\bar{p}_r \partial^i \bar{x}^r = -x^i \qquad\qquad \bar{x}^r \partial_i \bar{p}_r = -p_i .$$

By differentiating either the equations I_a or II_b one may verify the bracket relationships

(for I_a and II_b)
$$\{x^i, x^j\} = \partial_i \bar{p}_r \partial_j \bar{x}^r - \partial_j \bar{p}_r \partial_i \bar{x}^r = 0,$$
$$\{p_i, p_j\} = \partial^i \bar{p}_r \partial^j \bar{x}^r - \partial^j \bar{p}_r \partial^i \bar{x}^r = 0$$
$$\{p_i, x^j\} = \partial^i \bar{p}_r \partial_j \bar{x}^r - \partial_j \bar{p}_r \partial^i \bar{x}^r = \delta^i{}_j .$$

Similarly the transformations II_a or I_b yield

(for II_a and I_b) $\quad \{x^i, x^j\} = 0, \quad \{p_i, p_j\} = 0, \quad \{p_i, x^j\} = -\delta^i{}_j .$

Using a device originated by L. P. Eisenhart and M. S. Knebelman (1) we employ the bracket relations to identify the expressions

$$dx^i = \bar{\delta}_j x^i \, d\bar{x}^j + \bar{\delta}^j x^i \, d\bar{p}_j \qquad dp_i = \bar{\delta}_j p_i \, d\bar{x}^j + \bar{\delta}^j p_i \, d\bar{p}_j$$
$$dx^i = \partial^i \bar{p}_j \, d\bar{x}^j - \partial^i \bar{x}^j \, d\bar{p}_j \qquad dp_i = -\partial_i \bar{p}_j \, d\bar{x}^j + \partial_i \bar{x}^j \, d\bar{p}_j$$

with the subsequent results

(for I_a and II_b) $\quad \bar{\delta}_j x^i = \partial^i \bar{p}_j , \quad \bar{\delta}^j x^j = -\partial^i \bar{x}^j , \quad \bar{\delta}_j p_i = -\partial_i \bar{p}_j .$

In like manner we find

(for II_a and I_b) $\quad \bar{\delta}_j x^i = -\partial^i \bar{p}_j , \quad \bar{\delta}^j x^i = \partial^i \bar{x}^j , \quad \bar{\delta}_j p_i = \partial_i \bar{p}_j .$

These serve to establish the homogeneity properties

(for I_a and II_a)
$$\bar{x}^i(cx, kp) = c\bar{x}^i(x, p)$$
$$\bar{p}_i(cx, kp) = k\bar{p}_i(x, p)$$

(for I_b and II_b)
$$\bar{x}^i(cx, kp) = k\bar{x}^i(x, p)$$
$$\bar{p}_i(cx, kp) = c\bar{p}_i(x, p)$$

so that the transformations may be written

(I_a and II_a) $\bar{x}^i = x^r \partial_r \bar{x}^i , \ \bar{p}_i = p_r \partial^r \bar{p}_i$ (I_b and II_b) $\bar{x}^i = p_r \partial^r \bar{x}^i , \ \bar{p}_i = x^r \partial_r \bar{p}_i .$

It may be verified that the four elements I_a, II_a, I_b, II_b form a group with the identity I_a which is isomorphic with Klein's four-group.

17. A Complete Determination of a Factor Frame from an Affine Connection in a Flat Element Space

We consider in this section a flat element space in which there exist normal coordinates y^A for which the affine connection components $L^A{}_{BC}(y) = 0$. Since $\Omega^A{}_{BC} = 0$ the connection is symmetric and will be written as $\Gamma^A{}_{BC}$. Confining ourselves to the doubly homogeneous contact group, we purpose to define a factor frame $L^A{}_B$ in terms of $\Gamma^A{}_{BC}$ and the element vector (x^A). To this end we first observe that the invariant numerical tensor Π_{CA} has a vanishing covariant derivative in normal coordinates, and hence in all coordinates $\Pi_{QA}\Gamma^Q{}_{BC} + \Pi_{CQ}\Gamma^Q{}_{BA} = 0$, or

$$\begin{aligned} &\Gamma_{ijk} = \Gamma_{kji} \qquad \Gamma_{ij}{}^k = -\Gamma^k{}_{ji} \qquad \Gamma^i{}_j{}^k = \Gamma^k{}_j{}^i \\ &\Gamma^{ijk} = \Gamma^{kji} \qquad \Gamma_i{}^{jk} = -\Gamma^{kj}{}_i \qquad \Gamma_i{}^j{}_k = \Gamma_k{}^j{}_i . \end{aligned} \tag{17.1}$$

Since the covariant derivative of the vector x^A satisfies $x^A{}_{,B} - \delta^A{}_B = 0$ in normal coordinates we have the zero tensor $x^Q\Gamma^A{}_{BQ}$ whose expansion gives

$$p_r\Gamma^r{}_{ji} = x^r\Gamma_{rji}, \quad p_r\Gamma^{rji} = x^r\Gamma_r{}^{ji}, \quad p_r\Gamma^r{}_j{}^i = x^r\Gamma_{rj}{}^i, \quad p_r\Gamma^{rj}{}_i = x^r\Gamma_r{}^j{}_i \tag{17.2}$$

when reference is made to (17.1). We define the factor frame L_{ij} and L^{ij} by $L_{ij} = \Gamma_{ij} = p_r\Gamma^r{}_{ij}$ and $L^{ij} = \Gamma^{ij} = p_r\Gamma^{rij}$ so that $\Gamma_{ij} = \Gamma_{ji}$ and $\Gamma^{ij} = \Gamma^{ji}$ and shall demonstrate that the quantities Γ_{ij} and Γ^{ij} satisfy the conditions for a factor frame. First we note that the normal coordinates y^A for which $\Gamma^A{}_{BC}(y) = 0$ are likewise normal for $L^A{}_B$ since $L^A{}_B(y) = \delta^A{}_B$. Recalling the results of section 3, the transformation group now under discussion is the intersection group $\Sigma \frown \Sigma^*$ where Σ and Σ^* are characterized as all transformations from normal coordinates y^A to general coordinates x^A for which $|Y_ix^j| = \left|\frac{\partial x^j}{\partial y^i}\right| \neq 0$ and $|Y^jp_i| = \left|\frac{\partial p_i}{\partial q_j}\right| \neq 0$. But since $\frac{\partial x^j}{\partial y^i} = \frac{\partial p_i}{\partial q_j}$, Σ and Σ^* coincide. Furthermore, from $\delta^i{}_j = X_jy^rY_rx^j = Y^rp_jY_rx^i = \partial^rp_j\partial_rx^i = \partial^iq_r\partial_jy^r$ and from the bracket relationship $\partial^iq_r\partial_jy^r - \partial_jq_r\partial^iy^r = \delta^i{}_j$ it follows that the transformations $y \to x$ must satisfy

$$\partial_jq_r\partial^iy^r = 0. \tag{17.3}$$

If in the transformation law

$$\Gamma^A{}_{BC} = \bar\Gamma^Q{}_{RS}\bar\partial_Qx^A\partial_B\bar x^R\partial_C\bar x^S + \bar\partial_Qx^A\partial^2{}_{BC}\bar x^Q \tag{17.4}$$

we identify $\bar x^A$ with normal coordinates y^A, we have $\Gamma^A{}_{BC}(x) = \partial_Qx^A\partial^2{}_{BC}y^Q$, and on contracting with x^i, $x^r\Gamma_{rj}{}^k = x^r(\partial_sp_r\partial^2{}_j{}^k y^s + \partial^sp_r\partial^2{}_j{}^k q_s) = y^s\partial^2{}_j{}^k q_s$. But on differentiating $y^r\partial_sp_j = 0$, $y^s\partial^k(\partial_sp_j) + \partial^ky^s\partial_sp_j = -(y^s\partial^2{}_j{}^k q_s + \partial^ky^s\partial_jq_s) = 0$, and therefore

$$x^r\Gamma_{rj}{}^k = x^r\Gamma_r{}^k{}_j = 0. \tag{17.5}$$

To obtain the transformation laws of the frame components sum (17.4) in the form $\bar\Gamma^Q{}_{BR}\bar\partial_Qx^A\partial_C\bar x^R = \Gamma^A{}_{QC}\bar\partial_Bx^Q + \bar\partial^2{}_{BQ}x^A\partial_C\bar x^Q$ with the covariant vector $(x_A) =$

(p_i, x^j), $\bar{x}_Q \bar{\Gamma}^Q{}_{BE} \partial_C \bar{x}^E = x_Q \Gamma^Q{}_{EC} \bar{\partial}_B x^E + x_Q \partial_C (\bar{\partial}_B x^Q)$, and expand with reference to (17.2) and (17.5). The resulting equations fall into the three sets $\bar{L}_{jr} \bar{\partial}^r x^k = -L^{kr} \bar{\partial}_j p_r$, $\bar{L}_{jr} \partial_k \bar{x}^r = L_{rk} \bar{\partial}_j x^r + \partial_k \bar{p}_j$, $\bar{L}^{jr} \bar{\partial}_r x^k = L^{rk} \partial_r \bar{x}^j + \partial^k \bar{x}^j$ agreeing with the second line of (2.3) and with the form (11.2) of the first line of (2.2). This completes the identification of the quantities Γ_{ij} and Γ^{ij} as components of a factor frame.

18. A Real Decomposition Under the Complex Group

Let $(x^A) = (x^k, x^{\bar{k}})$ be $2q$ real variables in terms of which q complex variables z^k are defined by $z^k = x^k + ix^{\bar{k}}$. If z^k be subjected to an analytic transformation $\bar{z}^k = \bar{z}^k(z^j)$ the $2q$ real variables $(\bar{x}^A) = (\bar{x}^k, \bar{x}^{\bar{k}})$ defined by $\bar{z}^k = \bar{x}^k + i\bar{x}^{\bar{k}}$ will satisfy the conditions

$$\partial_j \bar{x}^k = \partial_{\bar{j}} \bar{x}^{\bar{k}} \qquad \partial_{\bar{j}} \bar{x}^k = -\partial_j \bar{x}^{\bar{k}}. \tag{18.1}$$

These generalized Cauchy-Riemann equations fully characterize the complex group $\bar{x}^A = \bar{x}^A(x)$ induced on the $2q$ real variables x^A by the group of all analytic transformations on the q complex variables z^k. The conditions (18.1) may be written more concisely as the transformation law $E^Q{}_B \partial_Q \bar{x}^A = E^A{}_Q \partial_B \bar{x}^Q$ of a simple mixed tensor $E^A{}_B$ with the components $\left\| \begin{matrix} E^i{}_j & E^i{}_{\bar{j}} \\ E^{\bar{i}}{}_j & E^{\bar{i}}{}_{\bar{j}} \end{matrix} \right\| = \left\| \begin{matrix} 0 & -\delta^i{}_j \\ \delta^{\bar{i}}{}_j & 0 \end{matrix} \right\|$ in all coordinate systems.

There is a marked similarity between the complex and contact transformation groups which will become increasingly evident as we proceed. Just as $L^{jk} = 0$ served to characterize extended point transformations as a subgroup of the contact group, so here from the laws $\bar{L}^k{}_r(\partial_j \bar{x}^r + L^i{}_j \partial_i \bar{x}^r) = \partial_j \bar{x}^{\bar{k}} + L^{\bar{i}}{}_j \partial_{\bar{i}} \bar{x}^{\bar{k}}$ and $\bar{L}^{\bar{i}}{}_{\bar{r}}(\partial_k \bar{x}^{\bar{r}} + L^s{}_k \partial_s \bar{x}^{\bar{r}}) = \partial_k \bar{x}^{\bar{i}} + L^s{}_k \partial_s \bar{x}^{\bar{i}}$ it may be seen on identifying x^A with a system of coordinates y^A for which

$$L^i{}_k(y) = 0 \qquad L^{\bar{k}}{}_j(y) = 0 \tag{$R_{18.2}$}$$

and $\bar{x}^A$ with a new set of coordinates $\bar{y}^A$ with the same properties that these completely characterize the subgroup of the complex group consisting of all separated transformations of the form $\bar{y}^k = \varphi^k(y^r)$ and $\bar{y}^{\bar{k}} = \varphi^{\bar{k}}(y^{\bar{r}})$. We accordingly define the L's by ($R_{18.2}$) in some arbitrary but definite coordinate system y^A which will be called normal. As immediate consequences of these definitions we have $L^i{}_{\bar{j}} L^{\bar{j}}{}_k = 0$.

$$L^i{}_r L^r{}_j = 0, \qquad L^{\bar{i}}{}_j + L^{\bar{i}}{}_{\bar{j}} = 0, \qquad X_j \bar{x}^{\bar{i}} = X_{\bar{j}} \bar{x}^{\bar{i}},$$

$$\| \epsilon^A{}_B \| = \| M^A{}_Q E^Q{}_R L^R{}_B \| = \left\| \begin{matrix} \epsilon^i{}_j & \epsilon^i{}_{\bar{j}} \\ \epsilon^{\bar{i}}{}_j & \epsilon^{\bar{i}}{}_{\bar{j}} \end{matrix} \right\| = \left\| \begin{matrix} 0 & -\delta^i{}_j \\ \delta^{\bar{i}}{}_j & 0 \end{matrix} \right\|.$$

We complete the choice of affine connection made in section 7 by stipulating that the covariant derivatives of the tensor $\epsilon^A{}_B$ shall vanish, the conditions being

$$\epsilon^i{}_{k,j} = X_j \epsilon^{\bar{i}}{}_k - \epsilon^{\bar{i}}{}_r \Gamma^r{}_{jk} + \epsilon^{\bar{r}}{}_k \lambda^{\bar{i}}{}_{j\bar{r}} = 0 \qquad \epsilon^{\bar{i}}{}_{k,j} = X_{\bar{j}} \epsilon^{\bar{i}}{}_k - \epsilon^{\bar{i}}{}_r \lambda^r{}_{jk} + \epsilon^{\bar{r}}{}_k \Gamma^{\bar{i}}{}_{\bar{j}\bar{r}} = 0.$$

We define $l^i{}_{jk}$ and $l^{\bar{i}}{}_{j\bar{k}}$ as the solutions

$$(R_{18.3}) \qquad l^i{}_{jk} = \lambda^{\bar{i}}{}_{j\bar{k}} \qquad l\,{}_{j\bar{k}} = \lambda^{i}{}_{jk}$$

of these equations. Because all components of the curvature tensor vanish for our present choice of connection the spaces S^{2q} is flat. From section 3 the subgroups Σ and Σ^* of the full complex group corresponding to our choice of frame consist of all transformations from normal coordinates y^A to general coordinates x^A satisfying

$$\left|\frac{\partial x^i}{\partial y^j} + L^i{}_j(y)\frac{\partial x^{\bar{i}}}{\partial y^{\bar{r}}}\right| = \left|\frac{\partial x^i}{\partial y^j}\right| \neq 0 \quad \text{and} \quad \left|\frac{\partial x^{\bar{i}}}{\partial y^j} + L^r{}_j(y)\frac{\partial x}{\partial y^r}\right| = \left|\frac{\partial x^{\bar{i}}}{\partial y^j}\right| \neq 0$$

so that $\Sigma = \Sigma^*$.

19. A Complex Decomposition Under the Complex Group

The transformation equations of the factor frame are satisfied by the quantities $L^A{}_B$ with the components $\begin{Vmatrix} L^i{}_j & L^{\bar{i}}{}_j \\ L^i{}_{\bar{j}} & L^{\bar{i}}{}_{\bar{j}} \end{Vmatrix} = \begin{Vmatrix} \delta^i{}_j & -i\delta^{\bar{i}}{}_j \\ -i\delta^i{}_{\bar{j}} & \delta^{\bar{i}}{}_{\bar{j}} \end{Vmatrix}$ in all coordinate systems, for they become $-i\partial_j\bar{x}^i - \partial_j x^i = \partial_j\bar{x}^i - i\partial_j\bar{x}^i$ and $-i\partial_j\bar{x}^i - \partial_j x^i = \partial_j\bar{x}^i - i\partial_j\bar{x}^i$ which hold in virtue of the equations of definition of the complex group. The reciprocal matrix $\| M^A{}_B \|$ is given by $\begin{Vmatrix} M^i{}_j & M^{\bar{i}}{}_j \\ M^i{}_{\bar{j}} & M^{\bar{i}}{}_{\bar{j}} \end{Vmatrix} = \begin{Vmatrix} \frac{1}{2}\delta^i{}_j & \frac{i}{2}\delta^{\bar{i}}{}_j \\ \frac{i}{2}\delta^i{}_{\bar{j}} & \frac{1}{2}\delta^{\bar{i}}{}_{\bar{j}} \end{Vmatrix}$.

The equations

$$T: \begin{matrix} z^i = x^i + ix^{\bar{i}} \\ z^{\bar{i}} = x^i - ix^{\bar{i}} \end{matrix} \qquad T^{-1}: \begin{matrix} x^i = \frac{1}{2}(z^i + z^{\bar{i}}) \\ x^{\bar{i}} = \frac{i}{2}(z^i - z^{\bar{i}}) \end{matrix}$$

represent a transformation from the $2q$ real variables x^A to the q complex variables z^i and their conjugates $z^{\bar{i}}$. The vector components Z and factor components λ of a real covariant vector Λ_A and of a real contravariant vector Λ^A in the new complex coordinates z^A will be given by

$$Z_k = \frac{1}{2}(\Lambda_k - i\Lambda_{\bar{k}}) = \frac{1}{2}\lambda_k, \qquad Z_{\bar{k}} = \frac{1}{2}(\Lambda_k + i\Lambda_{\bar{k}}) = \frac{i}{2}\lambda_{\bar{k}};$$

$$Z^k = \Lambda^k + i\Lambda^{\bar{k}} = 2\lambda^k, \qquad Z^{\bar{k}} = \Lambda^k - i\Lambda^{\bar{k}} = -2i\lambda^{\bar{k}}.$$

The following relations will be needed in our discussion and may be verified without difficulty:

$$X_k = 2\frac{\partial}{\partial z^k}, \qquad X_j \bar{x}^i = \frac{\partial \bar{z}^i}{\partial z^j}, \qquad X_{\bar{k}} = -2i\frac{\partial}{\partial \bar{z}^k},$$

(19.1)

$$X_{\bar{k}} \bar{x}^{\bar{i}} = \frac{\partial \bar{z}^{\bar{i}}}{\partial z^{\bar{k}}}, \qquad \widehat{\left(\frac{\partial \bar{z}^i}{\partial z^k}\right)} = \frac{\partial \bar{z}^{\bar{i}}}{\partial z^{\bar{k}}}, \qquad \widehat{\left(\frac{\partial^2 \bar{z}^i}{\partial z^j \partial z^k}\right)} = \frac{\partial^2 \bar{z}^{\bar{i}}}{\partial z^{\bar{j}} \partial z^{\bar{k}}}.$$

The symbol $\frown$ designates the complex conjugate. Under the transformation $\bar{z}^{\bar{k}} = \bar{z}^{\bar{k}}(z^{\bar{i}})$, $\bar{z}^{\bar{k}} = \bar{z}^{\bar{k}}(z^{\bar{i}})$, the z components Z_A and Z^A of Λ_A and Λ^A transform by

$$\bar{Z}_k = Z_r \frac{\partial z^r}{\partial \bar{z}^k}, \qquad \bar{Z}_{\bar{k}} = Z_{\bar{r}} \frac{\partial z^{\bar{r}}}{\partial \bar{z}^{\bar{k}}}; \qquad \bar{Z}^k = Z^r \frac{\partial \bar{z}^k}{\partial z^r}, \qquad \bar{Z}^{\bar{k}} = Z^{\bar{r}} \frac{\partial \bar{z}^{\bar{k}}}{\partial z^{\bar{r}}}$$

and by merely replacing Z_A with λ_A and Z^A with λ^A these same equations give the transformation laws of the factor components.

We now come to the selection of an affine connection. Let us first consider a general affine connection $L^A{}_{BC}(z^D)$. Because of the separated form of the transformations $z^A \to \bar{z}^A$ all components of affine connection will transform as complex tensors with the exception of the sets $L^i{}_{jk}$ and $L^{\bar{i}}{}_{j\bar{k}}$ which will obey

$$\bar{L}^r{}_{jk} \frac{\partial z^i}{\partial \bar{z}^r} = L^i{}_{rs} \frac{\partial z^r}{\partial \bar{z}^j} \frac{\partial z^s}{\partial \bar{z}^k} + \frac{\partial^2 z^i}{\partial \bar{z}^j \partial \bar{z}^k}, \qquad \bar{L}^{\bar{r}}{}_{j\bar{k}} \frac{\partial z^{\bar{i}}}{\partial \bar{z}^{\bar{r}}} = L^{\bar{i}}{}_{\bar{r}\bar{s}} \frac{\partial z^{\bar{r}}}{\partial \bar{z}^j} \frac{\partial z^{\bar{i}}}{\partial \bar{z}^k} + \frac{\partial^2 z^{\bar{i}}}{\partial \bar{z}^j \partial \bar{z}^k}.$$

From (19.1) it follows that $\widehat{(L^i{}_{jk})} - L^i{}_{j\bar{k}}$ is a tensor and therefore we make the invariant restrictions

$$(R_{19.2}) \qquad L^i{}_{j\bar{k}} = \widehat{(L^i{}_{jk})}.$$

In selecting factor components of affine connection $l^A{}_{BC}$ we retain the restrictions of section 7 leaving us with but two sets of components $l^i{}_{jk}$ and $l^{\bar{i}}{}_{j\bar{k}}$ transforming by

$$(5.3) \quad l'^i{}_{jk} \bar{X}_r x^i = l^i{}_{rs} \bar{X}_j x^r \bar{X}_k x^s + \bar{X}_j \bar{X}_k x^i \qquad l'^{\bar{i}}{}_{j\bar{k}} \bar{X}_{\bar{r}} x^{\bar{i}} = l^{\bar{i}}{}_{\bar{r}\bar{s}} \bar{X}_j x^{\bar{r}} \bar{X}_{\bar{k}} x^{\bar{i}} + \bar{X}_j \bar{X}_{\bar{k}} x^{\bar{i}}$$

yet to be defined. From (19.1) it follows that $l^i{}_{jk} - 2L^i{}_{jk}$ and $l^{\bar{i}}{}_{j\bar{k}} + 2iL^{\bar{i}}{}_{j\bar{k}}$ are tensors and therefore we define the only non-vanishing factor components of connection by

$$(R_{19.3}) \qquad l^i{}_{jk} = 2L^i{}_{jk} \qquad l^{\bar{i}}{}_{j\bar{k}} = -2iL^{\bar{i}}{}_{j\bar{k}}.$$

The formulas (7.3) for the covariant derivatives of factor vectors are now simply multiples of the covariant derivatives of the complex vectors Z_k, Z^k and their complex conjugates. The restriction $(R_{19.2})$ identifies the complex conjugate of the covariant derivative of a vector Z with the covariant derivative of the complex conjugate vector $\hat{Z}$.

The complex decomposition of this section is valid for the full complex group,

for the criteria for admissible transformations $x \to \bar{x}$ are $|X_j\bar{x}^i| = \left|\frac{\partial \bar{z}^i}{\partial z^j}\right| \neq 0$ and $|X_j\bar{x}^i| = \left|\widehat{\frac{\partial \bar{z}^i}{\partial z^j}}\right| \neq 0$.

PRINCETON UNIVERSITY

REFERENCES

(1) L. P. Eisenhart and M. S. Knebelman, "Invariant theory of homogeneous contact transformations", Annals of Math., 37 (1936), 747–765.

(2) J. A. Schouten, "Zur Differentialgeometrie der Gruppe der Berührungstransformationen. I. Doppelthomogenen Behandlung vor Berührungstransformationen," Proc. Kon. Akad. Wet., Amsterdam, vol. 40 (1937) p. 100.

(3) F. Engel and K. Faber, "Die Liesche Theorie der partiellen Differentialgleichungen erster Ordnung", B. G. Teubner, 1932, p. 38.

(4) L. P. Eisenhart, "Contact transformations", Annals of Math., 30 (1928), p. 235.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

REPRESENTATION OF ERGODIC FLOWS

BY WARREN AMBROSE

(Received December 28, 1940)

Introduction

The theory of measure preserving transformations and flows (a flow is a 1-parameter group of measure preserving transformations) originated in the study of classical dynamical systems. The fact that von Neumann's proof of the mean ergodic theorem and G. D. Birkhoff's proof of the ergodic theorem make no use of the strong regularity conditions fulfilled by classical dynamical systems together with the fact that these theorems have applications in other fields where such strong regularity conditions are not fulfilled (notably in probability theory) has however led to a study of measure preserving transformations in terms of purely measure-theoretic formulations.[1] The present paper is concerned with the theory of flows from this measure-theoretic standpoint.

The principle result of this paper (Theorem 2) is a theorem which asserts that every (measurable) ergodic flow (an ergodic flow is one in which every set of positive measure sweeps out the whole space) is isomorphic (with respect to all measure properties) to a certain kind of flow that we call a flow "built under a function."[2] This representation theorem has various consequences of which the most important probably is the following: In studying the flows of classical dynamics one is constantly picking out cross sections, drawing tubes about the trajectories and, generally speaking, considering separately what happens along and what happens perpendicular to the direction of flow. This isomorphism theorem makes possible a similar analysis,—for an ergodic flow,—[3] under the most general measure theoretic formulations. It makes possible such an analysis because for a flow built under a function there exist certain sets with the properties of a cross section and the measure on the space is the direct product measure of a certain measure on the cross section with Lebesgue measure along the trajectories; these are the essential features necessary for such an analysis. As in the classical cases various properties of the flow become intimately tied up with the properties of a related transformation on this cross section. We re-

[1] See for example III, IV and X in the bibliography at the end of this paper.

[2] Flows of this kind were first introduced in a special case, by von Neumann in X, p. 636.

[3] Recently Prof. S. Kakutani has succeeded in extending this theorem to non-ergodic flows, using Theorem 1 of this paper. His method is similar to that which we use for the ergodic case. This and some other similar theorems will appear in a joint paper by Prof. Kakutani and the present author under the title, "Structure and Continuity of Measurable Flows." I take this opportunity to express my indebtedness to Prof. Kakutani for very many very helpful discussions of all the matters with which this paper is concerned.

mark that this representation theorem also enables us to show (except for one trivial and highly uninteresting kind of situation) that the trajectory of any point is always a measurable set (this we prove even for non-ergodic flows).

In section 1 we give our definitions and in section 2 we prove our general representation theorem. In section 3 we discuss the group of unitary operators always (since Koopman, V) associated with a flow and establish a spectral condition that a flow have a special kind of representation as a flow built under a function. We conclude by extending a theorem of von Neumann about differentiable ergodic flows to measurable ergodic flows; this extension is possible just because our isomorphism theorem enables us to obtain in the general case the tubes and cross sections for whose existence von Neumann assumed differentiability.

1. Definitions

DEFINITION 1. A *measure space* is a space, Ω, on which a completed[4] (countably additive) measure, $m(M)$, is defined, for which $0 < m(\Omega) < \infty$ and such that there exists at least one measurable set, M, for which $0 < m(M) < m(\Omega)$.

DEFINITION 2. A *measure preserving transformation*, T, is a 1:1 point transformation of a measure space Ω_1 onto[5] a measure space Ω_2 with the property that $M_2 = TM_1$ is measurable (in Ω_2) if and only if M_1 is measurable (in Ω_1) and for which $m_1(M_1) = m_2(M_2)$ for every measurable set M_1.[6]

DEFINITION 3. A *flow* is a 1-parameter family, T_t, $(-\infty < t < \infty)$ of measure preserving transformations of a measure space onto itself, which has the group property: $T_tT_s = T_{t+s}$[7] for all t and s.

DEFINITION 4. Let T_t be a flow on Ω. The set $M \subset \Omega$ is *invariant* under T_t if whenever P is in M then T_tP is in M for all t.

DEFINITION 5. The flow T_t is *ergodic* if there are no measurable sets invariant under T_t except possibly sets of measure 0 and complements of sets of measure 0.

DEFINITION 6. Let T_t be a flow on Ω' and let S_t be a flow on Ω''. T_t and S_t are *isomorphic* if it is possible to split Ω' into two disjoint measurable sets whose sum is Ω', Ω_1' and Ω_2', and to split Ω'' into two disjoint measurable sets whose sum is Ω'', Ω_1'' and Ω_2'', in such a way that:

1) Each of Ω_1' and Ω_2' is invariant under T_t and each of Ω_1'' and Ω_2'' is invariant under S_t.

2) Both Ω_2' and Ω_2'' have measure 0.[8]

---

[4] The measure $m(M)$ is completed if every subset of each set of measure 0 is measurable.

[5] We use the term "onto" with the customary meaning, that every point of Ω_{22} is the image of some point in Ω_1.

[6] Here we use m_1 for measure on Ω_1 and m_2 for measure on Ω_2. Obviously T is a measure preserving transformation of Ω_1 onto Ω_2 if and only if T^{-1} is a measure preserving transformation of Ω_2 onto Ω_1. Usually Ω_1 and Ω_2 will be the same.

[7] If T and S are transformations then by "$T = S$" we mean that $TP = SP$ for *every* point P.

[8] We allow the possibility that one or both of Ω_2' and Ω_2'' is empty.

3) There exists a measure preserving transformation, R, of Ω_1' onto Ω_1'' such that $T_tP = R^{-1}S_tRP$ for all P in Ω_1' (and all t).

Throughout this paper L will denote the real line taken with Lebesgue measure. If Ω is a measure space then $\Omega \times L$ will denote the product space of Ω with L, taken with the completed direct product measure defined in terms of Lebesgue measure on L and the given measure on Ω.

DEFINITION 7. Let T_t be a flow on Ω. T_t is *measurable* if the function T_tP is a measurable (P, t)-function, i.e. if for every measurable set, M, in Ω the (P, t)-set for which T_tP is in M is measurable in $\Omega \times L$.[9]

Equivalent to this definition is the assertion that for every complex-valued measurable function $f(P)$ defined on Ω the function $f(T_tP)$ is a measurable (P, t)-function.

DEFINITION 8. Let Ω be a measure space and T a measure preserving transformation of Ω onto itself. Consider the product space $\Omega \times L$ of Ω and the real line L (where measure on this product space is defined multiplicatively in terms of Lebesgue measure on L and the given measure on Ω). Let $f(P)$ be a real valued integrable function defined on Ω with $f(P) > c > 0$ for all P in Ω, and let $\bar{\Omega}$ be the portion of $\Omega \times L$ under the graph of $f(P)$, i.e. let $\bar{\Omega}$ be the set of points $\bar{P} = (P, x)$ for which $0 \leqq x < f(P)$. Then $\bar{\Omega}$ is a measure space. Define the flow T_t on $\bar{\Omega}$ by

$$
(1.1)\quad \begin{cases} T_t(P, x) = (T^nP, t + x - f(P) - \cdots - f(T^{n-1}P)) \\ \qquad \text{for } n > 0 \text{ and } f(P) + \cdots + f(T^{n-1}P) - x \\ \qquad\qquad \leqq t < f(P) + \cdots + f(T^nP) - x, \\ T_t(P, x) = (P, t + x) \text{ for } -x \leqq t < f(P) - x, \\ T_t(P, x) = (T^{-n}P, t + x + f(T^{-1}P) + \cdots + f(T^{-n}P)) \\ \qquad \text{for } n > 0 \text{ and } -f(T^{-1}P) - \cdots - f(T^{-n}P) - x \\ \qquad\qquad \leqq t < -f(T^{-1}P) - \cdots - f(T^{-n+1}P) - x. \end{cases}
$$

We call T_t the flow *built on the measure preserving transformation T under the function $f(P)$*.[10]

In this definition we have demanded that $f(P)$ be integrable only to make $\bar{\Omega}$ have finite measure; if we were considering flows on spaces of infinite measure then we would only have demanded that $f(P)$ be measurable and > 0. We might have required that $f(P)$ be positive everywhere rather than uniformly positive; this distinction is unessential but with $f(P)$ uniformly positive some of our proofs are simpler.

---

[9] In case Ω is a topological space and the measure is "properly" related to the topology it is usual to define a measurable flow as one with the property that for every open set 0 in Ω the (P, t)-set for which $T_tP \in 0$ is measurable in $\Omega \times L$. This apparently less restrictive definition is however equivalent to ours. For a proof see IV, pp. 9, 10.

[10] We shall not prove that the T_t thus defined actually are measure preserving transformations. A proof that they carry measurable sets into measurable sets is contained in the proof of Theorem 1. An application of Fubini's theorem then shows that they are measure preserving.

In considering flows built under a function it will be essential to consider the functions $F(\bar{P})$ and $G(\bar{P})$ defined by

$$F(\bar{P}) = F(P, x) = f(P),^{11} \tag{1.2}$$

$$G(\bar{P}) = G(P, x) = x. \tag{1.3}$$

Intuitively, $G(\bar{P})$ is the length of time since a particle now at $\bar{P}$ was last in Ω, while $F(\bar{P})$ is the length of time since it was last in Ω plus the length of time before it will again be in Ω.

The sets defined by $G(\bar{P}) =$ constant give us cross sections for such a flow.

If M is any set in Ω we call the $\bar{P}$-set defined by

$$M^* = \underset{(P,x)}{[} P \in M]$$

a *tube* and we say that M^* is based on M. If M^* is a tube with $F(\bar{P}) \geqq d > 0$ for every $\bar{P}$ in M^* and if $0 \leqq a < b \leqq d$ we call the set $M^*(a, b)$[12] defined by

$$M^*(a, b) = M^* \cdot \underset{\bar{P}}{[} a \leqq G(\bar{P}) < b]$$

a *box*, and the set $M^*(b)$ defined by

$$M^*(b) = M^* \underset{\bar{P}}{[} G(\bar{P}) \geqq b]$$

a *leftover* of the tube. We say that M^* is the tube *through* the box $M^*(a, b)$.[13] Let M^* be a tube and let $\bar{M}_k$ and $\bar{M}_{k,j}$ be defined, for each positive integer n, by

$$\begin{aligned} \bar{M}_k &= \underset{\bar{P}}{[} k2^{-n} \leqq F(\bar{P}) < (k+1)2^{-n}] \cdot M^*, \qquad k = 0, 1, 2, \cdots \\ \bar{M}_{k,j} &= \underset{\bar{P}}{[} j2^{-n} \leqq G(\bar{P}) < (j+1)2^{-n}] \cdot \bar{M}_k \quad j = 0, 1, \cdots, k-1. \end{aligned} \tag{1.4}$$

It is readily verified that

$$\sum_{k=0}^{\infty} \sum_{j=0}^{k-1} \bar{M}_{k,j} \to \sum_{k=0}^{\infty} \bar{M}_k = M^* \qquad (n \to \infty). \tag{1.5}$$

This fact will frequently be useful in what follows because it shows that every tube is a limit of sums of boxes.

2. The first representation theorem

In our definition of a flow built under a function we specified that the measure on $\bar{\Omega}$ be the direct product measure of a measure on Ω with Lebesgue measure on the x-axis. Suppose, however, that we have a flow defined by (1.1) on a region $\bar{\Omega}$ without assuming that the measure on $\bar{\Omega}$ is such a direct product measure, and suppose that we do not even have a measure defined on Ω (so that,

[11] We shall use the notations $\bar{P}$ and (P, x) interchangeably for points in $\bar{\Omega}$.

[12] We shall adhere to this notation, M^* for a tube, and $M^*(a, b)$ for a box through M^*, throughout this paper.

[13] Obviously a box uniquely defines the tube through itself.

in particular, we are not assuming that $f(P)$ is a measurable function or that T is a measure preserving transformation). The following theorem shows that if the flow satisfies certain measurability conditions then it is possible to put a measure on Ω for which T will be a measure preserving transformation and $f(P)$ will be a measurable function and such that the given measure on $\bar{\Omega}$ is the direct product of this measure on Ω with Lebesgue measure on the x-axis.

THEOREM 1. *Let T_t be a flow defined by* (1.1) *on a region $\bar{\Omega}$, with a measure $\bar{m}(\bar{M})$ on $\bar{\Omega}$. If T_t is a measurable flow and if the functions $F(\bar{P})$ and $G(\bar{P})$ are both $\bar{m}$-measurable, then there exists a measure $m(M)$ on Ω for which $f(P)$ is a measurable function and T is a measure preserving transformation and such that $\bar{m}(\bar{M})$ is the completed direct product measure*[14] *of $m(M)$ on Ω with Lebesgue measure on the x-axis.*

PROOF: We first define a Borel field[15] $\mathfrak{M}$ of sets in Ω and put a measure, $m(M)$, on it. We define $\mathfrak{M}$ to be the collection of those sets, M, in Ω with the property that M^*, the tube based on M, is $\bar{m}$-measurable. Obviously, $\mathfrak{M}$ is a Borel field. It is trivial from the fact that $F(\bar{P})$ is $\bar{m}$-measurable that $f(P)$ is measurable with respect to $\mathfrak{M}$. We define $m(M)$, for $M \in \mathfrak{M}$ by

$$m(M) = (1/c)\bar{m}(M^*(0, c)).^{16}$$

We want now to show that $m(M)$ is a completed measure and that, with this measure on Ω, T is a measure preserving transformation.

Because $F(\bar{P})$ and $G(\bar{P})$ are $\bar{m}$-measurable we know that if a tube is $\bar{m}$-measurable then so are its boxes and leftovers. Now we show that if a box is $\bar{m}$-measurable so is the tube through it. If a box is $\bar{m}$-measurable then any sub-box which has the same tube through it is $\bar{m}$-measurable because any such sub-box is obviously the intersection of the original box with a transform (by some T_t) of the original box. Also, if a box $M^*(a, b)$ is $\bar{m}$-measurable then the box $M^*(0, b)$ is $\bar{m}$-measurable since the latter is obviously the sum of a finite number of transforms of the former. To show that the tube through the box $M^*(a, b)$ is $\bar{m}$-measurable we see from (1.4) and (1.5) that it is sufficient to show that (for large n) each $\bar{M}_{k,j}$ is $\bar{m}$-measurable; the $\bar{m}$-measurability of the $\bar{M}_{k,j}$ (for large n) follows from the fact that

$$\bar{M}_{k,j} = T_{j2^{-n}}(M^*(0, 2^{-n}) \cdot [k2^{-n} \leqq F(\bar{P}) < (k+1)2^{-n}]).$$

---

[14] The region $\bar{\Omega}$ is not a product space unless $f(P) =$ constant so it cannot be precisely correct to say that $\bar{m}(\bar{M})$ is a direct product measure. Precisely expressed what we mean is that if we define $m'(M')$ on $\Omega \times L$ as the completed direct product measure of $m(M)$ on Ω and Lebesgue measure on L then the following is true: if $\bar{M} \subset \bar{\Omega}$ then $\bar{M}$ is $\bar{M}$-measurable if and only if it is m'-measurable and if it is $\bar{M}$-measurable then $\bar{m}(\bar{M}) = m'(\bar{M})$.

[15] By a Borel field we mean a collection of sets closed under the operations of complementation and countable addition.

[16] The notation $M^*(a, b)$ was introduced above. The number c here and in the remainder of this proof is some fixed positive number with the property that $f(P) > c$ for all P in Ω. It is easily seen (and in fact is a direct consequence of a later part of this proof) that $m(M)$ is independent of the particular choice of c.

It follows readily from what we have just shown that $m(M)$ is a completed measure. We shall now show that T preserves m-measure. Let $M \in \mathfrak{M}$ and $N = TM$; we shall show that the $\bar{m}$-measurability of $M^*(0, c)$ implies the $\bar{m}$-measurability of $N^*(0, c)$ and that these two sets have the same $\bar{m}$-measure. This will prove that if $M \in \mathfrak{M}$ then $TM \in \mathfrak{M}$ and $m(M) = m(TM)$. A similar proof shows that T^{-1} also has these properties so this will complete the proof that T is m-measure preserving. We define, for each positive integer n, the sets

$$\bar{M}^k = M^* \cdot \underset{\bar{P}}{[k2^{-n} \leqq F(\bar{P}) < (k+1)2^{-n}]} \underset{\bar{P}}{[k2^{-n} - c \leqq G(\bar{P}) < k2^{-n}]}.$$

It is readily verified that

$$T_c \sum_{k=0}^{\infty} \bar{M}^k \to N^*(0, c)$$

$$\sum_{k=0}^{\infty} T_{c-k/2^n} \bar{M}^k \to M^*(0, c) \quad (n \to \infty).$$

[17]

Because M^* is $\bar{m}$-measurable it follows from this that $N^*(0, c)$ and hence N^* is $\bar{m}$-measurable. Because in each of these sums the summands are disjunct it follows that $\bar{m}(M^*(0, c)) = \bar{m}(N^*(0, c))$ and hence that T is m-measure preserving.

Now we consider the space $\Omega \times L$ and in this space the Borel field $\mathfrak{B}_*$ determined by all sets of the form $M \times (a, b)$ where $M \in \mathfrak{M}$.[18] Defining measure multiplicatively in $\Omega \times L$[19] we have a measure, $m'(B_*)$, defined for all $B_* \in \mathfrak{B}_*$. Completing $\mathfrak{B}_*$ with respect to this measure we obtain a Borel field $\mathfrak{M}_*$. Now we note that $\bar{\Omega} \in \mathfrak{B}_*$; this follows from (1.5) applied to the tube $M^* = \bar{\Omega}$. We define $\mathfrak{B}'$ to be the Borel field of sets of the form $B_*\bar{\Omega}$ where $B_* \in \mathfrak{B}_*$, and $\mathfrak{M}'$ to be the Borel field of sets of the form $M_*\bar{\Omega}$ where $M_* \in \mathfrak{M}_*$.[20] Let $\bar{\mathfrak{M}}$ denote the Borel field of $\bar{m}$-measurable sets. Then what we want to show is that $\mathfrak{M}' = \bar{\mathfrak{M}}$ and that for sets $\bar{M} \in \bar{\mathfrak{M}}$ we have $m'(\bar{M}) = \bar{m}(\bar{M})$.

We now define $\mathfrak{F}'$ to be the field[21] consisting of those sets which are finite sums of boxes and leftovers of $\bar{m}$-measurable tubes. Now the Borel field determined by $\mathfrak{F}'$ is $\mathfrak{B}'$.[22] To see this we note first that $\mathfrak{B}'$ is the Borel field deter-

---

[17] The sets $\bar{M}_k$ and $\bar{M}_{k,j}$ are defined in terms of a given tube which we here take to be the tube through $M^*(a, b)$.

[18] By the interval (a, b) we shall mean the interval taken with its left end point a but without its right end point b. The Borel field determined by a collection of sets is the smallest Borel field containing that collection.

[19] For a discussion of product measures in abstract spaces see VI.

[20] An equivalent definition of $\mathfrak{M}'$ would have been to say that it is the Borel field obtained by completing $\mathfrak{B}'$ with respect to m'-measure.

[21] By a field we mean a collection of sets closed under the operations of complementation and finite addition.

[22] For a discussion of fields, Borel fields, etc. see VI.

mined by all sets of the form $(M \times (a; b)) \cdot \bar{\Omega}$ where $M \in \mathfrak{M}$.[23] Now each such set is easily seen to be the sum of a box of some $\bar{m}$-measurable tube and a leftover of some $\bar{m}$-measurable tube; on the other hand, every $\bar{m}$-measurable box, every $\bar{m}$-measurable tube and every leftover of an $\bar{m}$-measurable tube is in $\mathfrak{B}'$, so $\mathfrak{F}' \subset \mathfrak{B}'$. This proves that $\mathfrak{B}'$ is the Borel field determined by $\mathfrak{F}'$.

Both $F(\bar{P})$ and $G(\bar{P})$ are measurable with respect to $\mathfrak{B}'$; this is true for $F(\bar{P})$ because the $\bar{P}$-set for which $F(\bar{P}) > b$ is, for any b, an $\bar{m}$-measurable tube and such a tube we already know to be in $\mathfrak{B}'$. This is true for $G(\bar{P})$ because

$$[G(\bar{P}) < b] = [F(\bar{P}) < b] + [F(\bar{P}) \geqq b]\,[G(\bar{P}) < b].$$

The first set on the right side of this equality is an $\bar{m}$-measurable tube and the second is an $\bar{m}$-measurable box, so $G(\bar{P})$ is measurable with respect to $\mathfrak{B}'$.

We shall now prove that $\mathfrak{M}' \subset \overline{\mathfrak{M}}$ and that if $\bar{M} \in \mathfrak{M}'$ then $m'(\bar{M}) = \bar{m}(\bar{M})$. We know that every $\bar{m}$-measurable box belongs to $\overline{\mathfrak{M}}$, and that every leftover of an $\bar{m}$-measurable tube belongs to $\overline{\mathfrak{M}}$; it follows that $\mathfrak{F}'$ and then $\mathfrak{B}'$ is included in $\overline{\mathfrak{M}}$. We shall show that if $\bar{M} \in \mathfrak{B}'$ then $m'(\bar{M}) = \bar{m}(\bar{M})$. This trivially implies that $\mathfrak{M}' \subset \overline{\mathfrak{M}}$ and that m'-measure and $\bar{m}$-measure agree on $\mathfrak{M}'$. Consider an $\bar{m}$-measurable box, $M^*(a, b)$; let n be a fixed positive integer and write

$$\begin{aligned} M^*(a, b) = {} & M^*(a, a + [b - a]/n) + M^*(a + [b - a]/n, a + 2[b - a]/n) \\ & + \cdots \\ & \cdots \\ & \cdots \\ & + M^*(b - [b - a]/n, b). \end{aligned}$$

These sets are all disjunct and are transforms of one another (under members of the group T_t); so each has $\bar{m}$-measure equal to $(1/n)\bar{m}(M^*(a, b))$. It follows readily that for y real and $a < y < b$ we have

$$\bar{m}(M^*(a, y)) = [(y - a)/(b - a)]\bar{m}(M^*(a, b)). \tag{2.1}$$

From the definition of m'-measure we have

$$m'(M^*(0, c)) = c(1/c)\bar{m}(M^*(0, c)) = \bar{m}(M^*(0, c)). \tag{2.2}$$

Taken together (2.1) and (2.2) imply that $m'(\bar{M}) = \bar{m}(\bar{M})$ whenever $\bar{M}$ is an $\bar{m}$-measurable box. By (1.5) it follows that this is true whenever $\bar{M}$ is an $\bar{m}$-measurable tube and hence for all $\bar{M}$ in $\mathfrak{F}'$. Hence $m'(\bar{M}) = \bar{m}(\bar{M})$ for all $\bar{M} \in \mathfrak{B}'$.[24] Hence $\mathfrak{M}' \subset \overline{\mathfrak{M}}$ and for sets in $\mathfrak{M}'$ the measures m' and $\bar{m}$ agree.

---

[23] This is a consequence of the following trivial lemma: If $\Omega_1 \supset \Omega_2$, $\{F_\alpha\}$ is a collection of sets in Ω_1, $\mathfrak{B}_1$ is the Borel field determined by the F_α, and $\Omega_2 \in \mathfrak{B}_1$ then the Borel field of sets of the form $M_1\Omega_2$ where $M_1 \in \mathfrak{B}_1$ is the same as the Borel field determined by sets of the form $F_\alpha\Omega_1$.

[24] The fact that if two (countably additive) measures agree on a field then they agree on the Borel field it determines is best proved by use of the theorem (VI, p. 85) that the normal family determined by a field is the same as the Borel field determined by the field.

Now we shall prove that each T_t takes sets of $\mathfrak{B}'$ into sets of $\mathfrak{B}'$ and sets of $\mathfrak{M}'$ into sets of $\mathfrak{M}'$. Because we know that $\mathfrak{M}' \subset \overline{\mathfrak{M}}$ and that the T_t are $\bar{m}$-measure preserving it is sufficient to prove the former, and because the T_t form a group it is sufficient to prove this for $0 \leqq t < c$. Because the $\bar{m}$-measurable boxes are a determining collection for $\mathfrak{B}'$ it is sufficient to show that any $\bar{m}$-measurable box goes into a set of $\mathfrak{B}'$. Let then $M^*(a, b)$ be an $\bar{m}$-measurable box and t a fixed real number, with $0 \leqq t < c$. Let M^* be the tube through $M^*(a, b)$, M the base of M^*, $N = TM$ and N^* the tube based on N. Then it is easily seen that

$$T_t M^*(a, b) = M^* \cdot \underset{\bar{P}}{[}a + t \leqq G(\bar{P}) < b + t]$$

$$+ \lim_{n \to \infty} \sum_{k=0}^{\infty} N^* \cdot \underset{\bar{P}}{[}k2^{-n} \leqq b + t - F(T_{-t}\bar{P}) < (k+1)2^{-n}]$$

$$\cdot \underset{\bar{P}}{[}k2^{-n} + a - b \leqq G(\bar{P}) < k2^{-n}].$$

Now M^* and N^* are $\bar{m}$-measurable tubes and hence in $\mathfrak{B}'$; the set for which $a + t \leqq G(\bar{P}) < b + t$ is in $\mathfrak{B}'$ because $G(\bar{P})$ is measurable with respect to $\mathfrak{B}'$, and the sets under the summation are $\bar{m}$-measurable boxes and hence are in $\mathfrak{B}'$. Consequently $T_t M^*(a, b) \in \mathfrak{B}'$.[25]

It remains to show that $\overline{\mathfrak{M}} \subset \mathfrak{M}'$. Define the set $\bar{H}$ by

$$\bar{H} = \underset{\bar{P}}{[}G(\bar{P}) < c].$$

To show that $\overline{\mathfrak{M}} \subset \mathfrak{M}'$ it will be sufficient to show that if $\bar{M} \in \overline{\mathfrak{M}}$ and $\bar{M} \subset \bar{H}$ then $\bar{M} \in \mathfrak{M}'$. This is sufficient because any set in $\overline{\mathfrak{M}}$ is obviously the sum of a countable number of transforms of sets in $\overline{\mathfrak{M}}$ which are included in $\bar{H}$ and we know that each T_t takes sets of $\mathfrak{M}'$ into sets of $\mathfrak{M}'$.

Now we consider the product space $\bar{\Omega} \times L$ (L is the real line and $\mathfrak{L}$ will denote the Lebesgue measurable sets) and in this product space four Borel fields: $\mathfrak{M}' \times \mathfrak{L}$, $\mathfrak{M}' \otimes \mathfrak{L}$, $\overline{\mathfrak{M}} \times \mathfrak{L}$, $\overline{\mathfrak{M}} \otimes \mathfrak{L}$. These are defined as follows: $M' \times \mathfrak{L}$ is the Borel field determined by sets of the form $M \times E$, where $M \in \mathfrak{M}'$ and $E \in \mathfrak{L}$; $\mathfrak{M}' \otimes \mathfrak{L}$ is the completion of this Borel field with respect to the multiplicative measure defined on it. $\overline{\mathfrak{M}} \times \mathfrak{L}$ and $\overline{\mathfrak{M}} \otimes \mathfrak{L}$ are defined similarly in terms of $\overline{\mathfrak{M}}$ and $\mathfrak{L}$. From the fact that $\mathfrak{M}' \subset \overline{\mathfrak{M}}$ it follows that $\mathfrak{M}' \otimes \mathfrak{L} \subset \overline{\mathfrak{M}} \otimes \mathfrak{L}$ and that if $\tilde{M} \in \mathfrak{M}' \otimes \mathfrak{L}$ then the measure defined for it as a set of $\mathfrak{M}' \otimes \mathfrak{L}$ is the same as the measure defined for it as a set of $\overline{\mathfrak{M}} \otimes \mathfrak{L}$. We shall denote this measure in $\bar{\Omega} \times L$ by $\tilde{m}(\tilde{M})$.

In $\bar{\Omega} \times L$ we are considering points (P, x, t). If $\bar{M}$ is a (P, x)-set and $\bar{M}_1$

[25] What we are proving here is that a flow defined by (1.1), with measure taken as a product measure, takes measurable sets into measurable sets.

is a (P, t)-set and if $\bar{M}$ contains the same pairs (P, y) as $\bar{M}_1$ we shall write $\bar{M} = \bar{M}_1$. If $\bar{M} \subset \bar{\Omega} \times L$ we shall use the following notations:

$$\bar{M}_{(P,x)} = [(P, x, t) \in \bar{M}]$$

$$\bar{M}^t = \underset{(P,x)}{[} (P, x, t) \in \bar{M}]$$

$$\bar{M}_x = \underset{(P,t)}{[} (P, x, t) \in \bar{M}].$$

Now we consider the transformation S of $\bar{\Omega} \times L$ onto itself defined by $S(P, x, t) = (P, x, t + x)$. This transformation takes sets of any one of our four Borel fields in $\bar{\Omega} \times L$ into sets of the same Borel field and preserves $\tilde{m}$-measure. We shall indicate the proof of this for $\overline{\mathfrak{M}} \times \mathfrak{L}$ and $\overline{\mathfrak{M}} \otimes \mathfrak{L}$, the proof for the other two being the same. First, if $\bar{M} = \bar{M} \times (a, b)$ then obviously $S\bar{M}$ is a limit of finite sums of such sets. Since S takes sums into sums and complements into complements it follows that S takes sets of $\overline{\mathfrak{M}} \times \mathfrak{L}$ into sets of $\overline{\mathfrak{M}} \times \mathfrak{L}$. The t-sets $\bar{M}_{(P,x)}$ and $(S\bar{M})_{(P,x)}$ are obviously congruent (for all (P, x)) and so have the same t-measure. Hence, by Fubini's theorem $\bar{M}$ and $S\bar{M}$ have the same $\tilde{m}$-measure. Because S preserves the measure of sets in $\overline{\mathfrak{M}} \times \mathfrak{L}$ it must take sets of $\overline{\mathfrak{M}} \otimes \mathfrak{L}$ into sets of $\overline{\mathfrak{M}} \otimes \mathfrak{L}$.

Now let $\bar{M} \in \overline{\mathfrak{M}}$, $\bar{M} \subset \bar{H}$, and consider the set $^*\bar{M}$ defined by

$$^*\bar{M} = (S \cdot \underset{(P,x,t)}{[}T_t(P, x) \in \bar{M}]) \left(\underset{(P,x,t)}{[} 0 \leqq t < c] \cdot \underset{(P,x,t)}{[} 0 \leqq x < c] \right).$$

Because T_t is a measurable flow and S is a measure preserving transformation in $\bar{\Omega} \times L$, $^*\bar{M} \in \overline{\mathfrak{M}} \otimes \mathfrak{L}$. We shall call this $^*\bar{M}$ the *associated set* of $\bar{M}$. It has the following readily established properties:

(2.3) for $0 \leqq x < c$: $\quad ^*\bar{M}_x = \bar{M}$

(2.4) for $0 \leqq t < c$: $\quad ^*\bar{M}^t = \bar{H} \underset{(P,x)}{[} P \in \bar{M}_t]$.[26]

We now prove a sequence of assertions which will show that $\overline{\mathfrak{M}} \subset \mathfrak{M}'$. We number them for convenience.

1) If $\bar{M} \in \overline{\mathfrak{M}}$ and $^*\bar{M}$ is its associated set then for almost all t, $^*\bar{M}^t$ is in $\mathfrak{M}'$ and for almost all x, $\bar{M}_x$ is in $\mathfrak{M}$.

Let $\bar{M} \in \overline{\mathfrak{M}}$; then for almost all t, $^*\bar{M}^t$ is $\bar{m}$-measurable (by Fubini's theorem) and for all t it is a box; hence $^*\bar{M}^t \in \mathfrak{M}'$ for almost all t. If t is such that $^*\bar{M}^t \in \mathfrak{M}'$ then it follows from (2.4) that $\bar{M}_t \in \mathfrak{M}$.[26] Hence $\bar{M}_x \in \mathfrak{M}$ for almost all x.

We now define $\bar{\mathfrak{M}}$ to be the collection of all sets $\bar{M}$ included in

$$\bar{H} = \underset{(P,x,t)}{[} 0 \leqq x < c] \cdot \underset{(P,x,t)}{[} 0 \leqq t < c].$$

[26] $\bar{M}_t$ denotes the P-set: $\underset{P}{[}(P, t) \in \bar{M}]$.

and belonging to $\overline{\mathfrak{M}} \otimes \mathfrak{L}$ which have the property that for almost all x, $\tilde{M}_x \in \mathfrak{M}'$.

2) If $\tilde{M} \subset \tilde{H}$ and $\tilde{M} \in \overline{\mathfrak{M}} \times \mathfrak{L}$ then $\tilde{M} \in \tilde{\mathfrak{M}}$.

To prove this it is sufficient,—since $\tilde{\mathfrak{M}}$ is obviously a Borel field,[27]—to show that $\tilde{\mathfrak{M}}$ contains all sets of the form $\bar{M} \times (a, b)$ where $\bar{M} \in \overline{\mathfrak{M}}$, so consider such a set. We know from 1) that for almost all x, $\bar{M}_x$ belongs to $\mathfrak{M}$. Now $(\bar{M} \times (a, b))_x = \overline{\mathfrak{M}}_x \times (a, b)$ which is in $\mathfrak{M}'$ for almost all x.

3) If $\bar{M} \in \overline{\mathfrak{M}}$ (and $\bar{M} \subset \bar{H}$) then

$$\tilde{m}(\bar{M}) = \int_0^c \bar{m}(\bar{M}^t)\, dt = c \int_0^c m(\bar{M}_x)\, dx.$$

This is a consequence of Fubini's theorem and the fact that $\bar{m}(\bar{M}^t) = cm(\bar{M}_t)$.

4) If $\bar{N} \in \overline{\mathfrak{M}}$ (and $\bar{N} \subset \bar{H}$) and has $\bar{m}$-measure 0 then $m(\bar{N}_x) = 0$ for almost all x in $(0, c)$.

If $\bar{N}$ has $\bar{m}$-measure 0 then

$$[\underset{(P,x,t)}{} T_t(P, x) \in \bar{N}]$$

has $\tilde{m}$-measure 0 (by Fubini's theorem and the fact that T_t is measurable). Hence the associated set, $*\bar{N}$, of $\bar{N}$ has $\tilde{m}$-measure 0. Then from

$$0 = \tilde{m}(*\bar{N}) = c \int_0^c m(\bar{N}_x)\, dx,$$

we see that $\bar{N}_x$ has m-measure 0 for almost all x in $(0, c)$.

5) If $\tilde{M} \in \tilde{\mathfrak{M}}$ (and $\tilde{M} \subset \tilde{H}$) and $\tilde{m}(\tilde{M}) = 0$ then $\bar{m}(\tilde{M}_x) = 0$ for almost all x in $(0, c)$.

By definition we know that $\tilde{M}_x \in \mathfrak{M}'$ except for an x-set (which we call θ_1) of Lebesgue measure 0. Also $\tilde{M}_{(P,x)}$ has t-measure 0 except for a (P, x)-set in $\overline{\mathfrak{M}}$ of $\bar{m}$-measure 0. Denote this exceptional (P, x)-set by $\bar{N}$. Then, by 4), except for an x-set (which we call θ_2) of Lebesgue measure 0 we have $m(\bar{N}_x) = 0$. We shall now show that if $x \notin \theta_1 + \theta_2$ then $\bar{m}(\tilde{M}_x) = 0$. Consider such an x. Then $\tilde{M}_x \in \mathfrak{M}'$. Then $\bar{N}_x$ has m-measure 0 and for P outside $\bar{N}_x$ we know that $m(\tilde{M}_{(P,x)}) = 0$. Hence, applying Fubini's theorem to the set $\tilde{M}_x$ (using the fact that it is in $\mathfrak{M}'$ and has t-measure 0 for almost all P) we see that $\bar{m}(\tilde{M}_x) = 0$.

6) If $\tilde{M} \subset \tilde{H}$ and $\tilde{M} \in \overline{\mathfrak{M}} \otimes \mathfrak{L}$ then $\tilde{M} \in \tilde{\mathfrak{M}}$.

We already (by 2)) know this to be true when $\tilde{M} \in \overline{\mathfrak{M}} \times \mathfrak{L}$, so it will be sufficient to show that if $\tilde{M} \in \overline{\mathfrak{M}} \times \mathfrak{L}$ and $\tilde{m}(\tilde{M}) = 0$ and if $\tilde{N} \subset \tilde{M}$ then $\tilde{N} \in \tilde{\mathfrak{M}}$. This is true because we know by 2) and 5) that for such an $\tilde{M}$ we have $\tilde{M}_x \in \mathfrak{M}'$ and $\bar{m}(\tilde{M}_x) = 0$ for almost all x, which implies that $\tilde{N}_x \in \mathfrak{M}'$ for almost all x.

We can now conclude our proof. Let $\bar{M} \in \overline{\mathfrak{M}}$, $\bar{M} \subset \bar{H}$. Then the associated set $*\bar{M} \in \overline{\mathfrak{M}} \otimes \mathfrak{L}$ and $*\bar{M} \subset \tilde{H}$; hence by 6) $\bar{M} \in \overline{\mathfrak{M}}$, so $*\bar{M}_x \in \mathfrak{M}'$ for almost all x in $(0, c)$. Since $*\bar{M}_x = \bar{M}$ for all x in $(0, c)$ this means that $\bar{M} \in \mathfrak{M}'$. Thus $\overline{\mathfrak{M}} \subset \mathfrak{M}'$ and hence $\overline{\mathfrak{M}} = \mathfrak{M}'$.

[27] I.e. $\tilde{\mathfrak{M}}$ is a Borel field when only sets within $\tilde{H}$ are considered and complements are taken with respect to $\tilde{H}$.

THEOREM 2. *Every measurable ergodic flow is isomorphic to a flow built under a function.*

PROOF: Let S_t be the flow, Ω^* the space on which it occurs, and $m^*(M^*)$ the measure on Ω^*. We want to find three things,—a space Ω, a transformation T, and a function $f(P)$,—such that the flow T_t defined in terms of them by (1.1) is isomorphic to the original flow. Then we shall be able to apply Theorem 1 to conclude that T_t is built under a function, i.e. that the measure involved is a direct product measure. The Ω we choose will be a certain subset of Ω^* obtained by a judicious picking of sequences of points along the trajectories of S_t, the transformation T on Ω will be the one taking each point of Ω into the next point along the trajectory (under S_t) which lies in Ω, and the $f(P)$ will be the length of time (measured in terms of S_t) it takes to move (under S_t) from P to TP.

Let M^* be an m-measurable set with $0 < m^*(M^*) < m^*(\Omega^*)$ and let $\varphi(P^*)$ be its characteristic function. For P^* fixed and outside a certain invariant (under S_t) set N_1^*, of m^*-measure 0 $\varphi(S_tP^*)$ is a measurable t-function. By a theorem of Wiener[28] we know that

$$\frac{1}{\epsilon}\int_0^\epsilon \varphi(S_tP^*)\,dt \to \varphi(P^*) \qquad (\epsilon \to 0),$$

for almost all P^*. Hence (applying Egoroff's theorem) it is possible to choose a number, a, with $0 \leqq a < 1$, such that for $\Phi(P^*)$ defined by

$$\Phi(P^*) = \frac{1}{a}\int_0^a \varphi(S_tP^*)\,dt$$ [29]

the P^*-sets

$$M_1^* = \underset{P^*}{[}\Phi(P^*) < \tfrac{1}{4}], \qquad M_2^* = \underset{P^*}{[}\Phi(P^*) > \tfrac{3}{4}]$$

both have positive m^*-measure. Also, for $P^* \notin N_1^*$ the t-function $\Phi(S_tP^*)$ is continuous. In fact it is obvious that, for $P^* \notin N_1^*$,

$$|\Phi(S_tP^*) - \Phi(S_sP^*)| \leqq (2/a)\,|t - s|. \tag{2.5}$$

Applying the ergodic theorem we know that for P^* fixed and outside a certain invariant set N_2^* of m^*-measure 0 the trajectory S_tP^* will have points in common with each of M_1^* and M_2^* for arbitrarily large and negatively arbitrarily large t. Now we write $\Omega_1^* = \Omega^* - [N_1^* + N_2^*]$ and $\Omega_2^* = N_1^* + N_2^*$; then Ω_1^* and Ω_2^* are m^*-measurable invariant sets and $m^*(\Omega_2^*) = 0$. We define Ω by

$$\Omega = \Omega_1^* \cdot \underset{P^*}{[}\Phi(P^*) = \tfrac{1}{2} \quad \text{and} \quad \Phi(S_tP^*) > \tfrac{1}{2} \quad \text{for all } t \text{ in } 0 < t \leqq a/8].$$

For the moment we shall not prove that Ω is an m^*-measurable set; this follows from the fact, proved below, that $G(\bar{P})$ is an $\bar{m}$-measurable function.

---

[28] Wiener, XI, Theorem III′, p. 2.

[29] $\Phi(P^*)$ is obviously m^*-measurable. Throughout this proof a is a fixed number.

Now we show that for any P^* in Ω_1^* the trajectory S_tP^* has points in common with Ω for arbitrarily large t; the same proof shows this to be true also for arbitrarily negatively large t. If t_0 is any positive number then there exist t_1 and t_2 such that $t_0 < t_1 < t_2$ and $S_{t_1}P^* \in M_1^*$, $S_{t_2}P^* \in M_2^*$. Because $\Phi(S_tP^*)$ is continuous in t there exists a t in (t_1, t_2) for which $\Phi(S_tP^*) = \frac{1}{2}$. We choose t' to be the greatest such t in (t_1, t_2). Then $S_{t'}P^* \in \Omega$ because $\Phi(S_{t'}P^*) = \frac{1}{2}$ while (2.5) implies

$$|t_2 - t'| \geqq (a/2)\,|\Phi(S_{t_2}P^*) - \Phi(S_{t'}P^*)| > a/8,$$

and hence that $\Phi(S_tS_{t'}P^*) > \frac{1}{2}$ for all t in $0 < t \leqq a/8$.

We define T and $f(P)$ as follows: if $P \in \Omega$ then there is a smallest positive number, t_1, for which $S_{t_1}P \in \Omega$.[30] Then $TP = S_{t_1}P$ and $f(P) = t_1$. We note that $f(P) > a/8$ for all P in Ω.

Now consider the region $\bar{\Omega}$ under the graph of $f(P)$[31], i.e. the set of points (P, x) for which $0 \leqq x < f(P)$, and let T_t be defined on $\bar{\Omega}$ by (1.1). We establish a 1:1correspondence, R, between Ω_1^* and $\bar{\Omega}$ as follows: if $(P, x) \in \bar{\Omega}$ then $P^* = S_xP$ is its corresponding point in Ω_1^*. This is obviously a 1:1 correspondence which carries S_t into T_t, i.e. if $P^* = R(P, x)$ then $S_tP^* = RT_t(P, x)$ for all t. In other words, $S_t\bar{P} = R^{-1}T_tR\bar{P}$ for all $\bar{P}$ in $\bar{\Omega}$. Now let $\bar{m}(\bar{M})$ be the measure on $\bar{\Omega}$ carried over from Ω_1^* by R.[32] If we knew $\bar{m}$-measure to be the direct product measure of a measure on Ω (for which T was a measure preserving transformation and $f(P)$ a measurable function) with Lebesgue measure on the x-axis then our theorem would be proved. To show this it is sufficient, by Theorem 1, to show that the functions $F(\bar{P})$ and $G(\bar{P})$ are $\bar{m}$-measurable.

To show that $G(\bar{P})$ is m-measurable let b be any number $\geqq 0$; then

$$R\{\bar{P}\,|\,G(\bar{P}) \leqq b\} = \sum_{0 \leqq t \leqq b} S_t\Omega,$$

so we must show the set on the right side to be m^*-measurable. This is shown by the following equalities (in these formulae some of the indices range over a whole interval; we adopt the following convention: when such a sum or a product is primed then the summation or product is to be taken over all rational numbers in the interval plus the end points of the interval):

$$\sum_{0 \leqq t \leqq b} S_t\Omega$$

$$= \sum_{0 \leqq t \leqq b} S_t\{[\Phi(P^*) = \tfrac{1}{2}][\Phi(S_sP^*) > \tfrac{1}{2} \quad \text{for } 0 < s \leqq a/8]\}$$

---

[30] That there exists a positive number t_1 with $S_{t_1}P \in \Omega$ follows from the last paragraph; there will be a smallest one (and it will be $> (a/8)$) because if $P \in \Omega$ then none of the points S_tP, for $0 < t \leqq (a/8)$, are in Ω.

[31] In this discussion we sometimes consider Ω as a subspace of Ω^* and sometimes as a space by itself.

[32] I.e. a set $\bar{M}$ is $\bar{m}$-measurable if and only if its correspondent in Ω_1 is m^*-measurable and its measure, $\bar{m}(\bar{M})$, is the m^*-measure of that correspondent.

$$= \sum_{0\leqq t\leqq b} \{[\Phi(S_{-t}P^*) = \tfrac{1}{2}][\Phi(S_{s-t}P^*) > \tfrac{1}{2} \quad \text{for } 0 < s \leqq a/8]\}$$

$$= \prod_{n=1}^{\infty} \sum_{0\leqq t\leqq b}' \{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n][\Phi(S_{s-t}P^*) > \tfrac{1}{2} \quad \text{for } 0 < s \leqq a/8]\}$$

$$= \prod_{n=1}^{\infty} \sum_{0\leqq t\leqq b}' \left\{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n] \prod_{k=1}^{\infty} [\Phi(S_{s-t}P^*) > \tfrac{1}{2} \quad \text{for } 1/k \leqq s \leqq a/8]\right\}$$

$$= \prod_{n=1}^{\infty} \sum_{0\leqq t\leqq b}' \left\{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n] \prod_{k=1}^{\infty} \sum_{m=1}^{\infty} [\Phi(S_{s-t}P^*) \geqq \tfrac{1}{2} + 1/m \right.$$

$$\left. \text{for } 1/k \leqq s \leqq a/8]\right\}$$

$$= \prod_{n=1}^{\infty} \sum_{0\leqq t\leqq b}' \left\{[|\Phi(S_{-t}P^*) - \tfrac{1}{2}| < 1/n] \prod_{k=1}^{\infty} \sum_{m=1}^{\infty} \prod_{1/k\leqq s\leqq a/8}' [\Phi(S_{s-t}P^*) \geqq \tfrac{1}{2} + 1/m]\right\},$$

which is obviously an m^*-measurable set.

To see that $F(\bar{P})$ is $\bar{m}$-measurable we note that

$$[F(\bar{P}) > b] = \sum_{0\leqq t\leqq b}' T_{-t}[G(\bar{P}) \geqq b].$$

We remarked in the introduction that Theorem 2 implies that, except for a trivial kind of exception, the trajectories in a measurable flow are always measurable sets. We shall now prove this. If S_t is any measurable flow on a space Ω^* then we can subdivide Ω^* into disjoint m^*-measurable invariant sets:

$$\Omega^* = \Omega_0^* + \Omega_1^* + \cdots + \Omega_n^* + \cdots,$$

(it may be that a finite or an infinite number of the Ω_n^* are empty; in fact for an ergodic flow all but one will be empty) in such a way that: 1) if $n \geqq 1$ then either Ω_n^* is empty or else $m^*(\Omega_n^*) > 0$ and S_t is ergodic on Ω_n^*, 2) either Ω_0^* is empty or else S_t is completely non-ergodic on Ω_0^*, i.e. if M^* is any measurable invariant set of positive measure then M^* contains an invariant set, N^*, for which $m^*(M^*) > m^*(N^*) > 0$. We see that every trajectory lying in Ω_0^* is measurable as follows: for each $\epsilon > 0$ it is clear that Ω_0^* can be divided into a finite number of disjoint measurable invariant sets (whose sum is Ω_0^*) each of measure $< \epsilon$. Each trajectory in Ω_0^* will lie in one of these sets. Hence each trajectory in Ω_0^* lies in some set of measure $< \epsilon$, and this for arbitrary positive ϵ. Consequently each trajectory in Ω_0^* lies in a set of measure 0 and therefore (since we always assume our measures to be completed) each trajectory in Ω_0^* is a measurable set of measure 0.

Since on each Ω_n^* (for $n \geqq 1$) which is non-empty S_t is ergodic (and since we took care of Ω_0^* in the last paragraph) the matter of the measurability of the trajectories will be cleared up completely if we can clear it up for ergodic flows. By Theorem 2 it will be sufficient (in considering ergodic flows) to consider flows built under a function. Let S_t be such a flow, on a region $\bar{\Omega}$, and let Ω, T, $f(P)$, $\bar{m}(\bar{M})$, $m(M)$ have their usual meanings. Clearly the ergodicity of S_t implies that

T is ergodic on Ω. We now distinguish three cases: 1) every point in Ω is m-measurable and of m-measure 0, 2) every point in Ω is m-measurable but some point has positive measure, 3) some point in Ω fails to be m-measurable. First we consider case 1) (the usual case). Let $S_t\overline{P}$ be any trajectory; this trajectory intersects Ω in a sequence of points $\overline{P}_n = (P_n, 0)$. Then this trajectory is the set

$$\sum_{n=-\infty}^{\infty} \underset{(P,x)}{[} P = P_n]$$

which is $\bar{m}$-measurable and of $\bar{m}$-measure 0, since our measure is a direct product measure. In case 2) the ergodicity of T implies that Ω consists of a finite number of points of equal positive m-measure plus perhaps a set of measure 0. Then S_t is obviously isomorphic to the well known periodic flow on the circle ($S_t x = x + t \mod \alpha$, $\alpha > 0$). In this case all the trajectories are $\bar{m}$-measurable sets but a certain one of them has positive measure while all others have measure 0. In case 3) it is clear (since m-measure is completed) that the m-measurable sets will contain a "chunk," i.e. an m-measurable set, M, of positive m-measure and such that if N is m-measurable and contained in M then either $m(N) = m(M)$ or $m(N) = 0$. The ergodicity of T implies that M has a finite number of transforms and that every m-measurable set of positive m-measure differs by at most a set of m-measure 0 from some finite sum of transforms of M. Then our flow is essentially the same as in case 2) except that instead of all the measure being concentrated on the trajectory swept out by a single point all the measure is concentrated on the set swept out by M; clearly every $\bar{m}$-measurable set must be (to within a set of $\bar{m}$-measure 0) of the form:

$$\sum_{t \in E} S_t M,$$

where E is any Lebesgue measurable t-set on some finite interval $(0, \alpha)$. (Obviously in this case it is possible to set up a measure preserving *set*-correspondence between this and a flow of case 2) so in a certain sense this is the same as case 2)). In this case there will be at least one non-measurable trajectory inside the set swept out by M. This is the only case in which there can be any non-measurable trajectories and this case is obviously of not the slightest interest.

To summarize the results of this discussion we may say that if T_t is any measurable flow then the space on which it occurs can be split into a completely non-ergodic part, in which all trajectories have measure 0, plus a denumerable number of ergodic parts (where a finite or infinite number of these parts may be empty) and that on each ergodic part there are three possibilities: 1) each trajectory may be measurable and of measure 0, 2) each trajectory may be measurable but with one of the trajectories (necessarily a periodic one) having positive measure, in which case the flow is isomorphic to a rotation on the circle, 3) all the measure may be concentrated on the set swept out by a single set, M, of measure 0 in such a way that the only $\bar{m}$-measurable sets are sums of transforms of M; in this case non-measurable trajectories are possible, but it is possible

to establish a measure preserving *set* correspondence between such a flow and a rotation on the circle.

3. The group U_t and a second representation theorem

Let S_t be a flow on a space Ω^* and let L_2 be the space of (complex-valued) functions $g(P^*)$ of integrable square on Ω^*. We define the operator U_t (for all real t) by

$$U_t g(P^*) = g(S_t P^*).^{33} \tag{3.1}$$

Obviously the U_t are unitary operators and form a 1-parameter group: $U_t U_s = U_{t+s}$. It is known that if U_t is a continuous t-function (in the sense that, for any g in L_2, $\| U_t g - g \| \to 0$ as $t \to 0$) then U_t has a spectral resolution, i.e. there exists a unique spectral family E_λ for which

$$U_t = \int_{-\infty}^{\infty} e^{it\lambda}\, dE_\lambda,^{34}$$

and it has been shown by Doob (III, Theorem 7) that if U_t is defined by (3.1), where S_t is a measurable flow, then U_t is a continuous t-function.[35] Hence U_t will have a spectral resolution in all cases we are considering.

We now define a special kind of flow built under a function and show that a measurable ergodic flow is isomorphic to a flow of this kind if and only if the group U_t has an eigenfunction of eigenvalue $\neq 0$. This result was proved,—except for the measure theoretic considerations,—by G. D. Birkhoff (II, p. 144) and our proof consists essentially in using his argument and then applying Theorem 2 to take care of these measure-theoretic considerations.

DEFINITION 9. An *eigenfunction* of the group U_t is a function $\psi(P^*)$ in L_2 which is different from 0 on some set of positive measure and such that for some real number λ,

$$U_t \psi = e^{i\lambda t} \psi$$

(for all real t). The number λ is called the *eigenvalue* of ψ.

DEFINITION 10. If a flow is built under a function $f(P)$, and $f(P)$ is a constant function[36] then we say the flow is *well built*. The number $f(P)$ is called the *height* of the flow.

---

[33] The introduction of this group of unitary operators and the idea of attempting to characterize the properties of a flow in terms of properties of the spectral resolution of this group of operators is due to B. O. Koopman, V.

[34] See VII and IX.

[35] If S_t is a measurable flow and U_t is defined by (3.1) then (by Fubini's theorem) $(U_t f, g)$ is a measurable t-function for each f and g in L_2 and Stone (VII) and von Neumann (IX) have shown that in case L_2 is separable this implies that U_t is a continuous t-function. In our case, however, where no separability conditions are assumed, it is necessary to use the deeper result of Doob to prove that U_t is a continuous t-function. We remark that for a measurable *ergodic* flow this continuity follows easily from Theorem 2.

[36] By constant we mean constant *everywhere*.

THEOREM 3. *A measurable ergodic flow is isomorphic to a well built flow if and only if its corresponding U_t group has an eigenfunction of eigenvalue $\neq 0$. The heights of the well built flows to which it is isomorphic are the numbers $2\pi/|\lambda|$, where λ is an eigenvalue of the U_t.*

PROOF: If T_t is a well built flow of height α and if λ is defined by $2\pi/\lambda = \alpha$ then obviously the function $\psi(P, x)$ defined by

$$\psi(P, x) = e^{i\lambda x}$$

is an eigenfunction of eigenvalue λ for the corresponding U_t. Since the eigenvalues form (for an ergodic flow) an additive group of real numbers[37] $-\lambda$ is also an eigenvalue. Hence if a measurable ergodic flow is isomorphic to a well built flow of height $2\pi/|\lambda|$ then both λ and $-\lambda$ are eigenvalues of the corresponding group U_t.

Now let S_t be a measurable ergodic flow on a space Ω^* and let the corresponding U_t have an eigenfunction ψ of eigenvalue λ, ($\lambda \neq 0$). Then $\bar{\psi}$ is an eigenfunction of eigenvalue $-\lambda$.[32] One of λ and $-\lambda$ is positive; we suppose λ is positive and from now on work with λ and ψ (if $-\lambda$ were the positive one we would work in the same way, but with $-\lambda$ and $\bar{\psi}$). Because ψ is an eigenfunction we know that for each t we have

$$\psi(S_t P^*) = e^{i\lambda t}\psi(P^*),$$

for almost all P^*. By changing $\psi(P^*)$ at most on a P^*-set of measure 0 we can make this equality hold for *all* P^* and t;[38] we suppose ψ to be so altered. Now $|\psi|$ is an invariant function (i.e. $|\psi(S_t P^*)| = |\psi(P^*)|$) and hence (since S_t is ergodic) $|\psi(P^*)| = \text{constant} = k > 0$ except for P^* in some m^*-measurable set, N^*, of m^*-measure 0.[39] Because $\psi(S_t P^*) = e^{i\lambda t}\psi(P^*)$ for all P^* and t we see that N^* is an invariant set. Now write $\Omega_2^* = N^*$ and $\Omega_1^* = \Omega^* - N^*$. Then define Ω by

$$\Omega = \Omega_1^* \cdot \underset{P^*}{[\psi(P^*) = k]}$$

(where k is the constant mentioned above). We establish a 1:1 correspondence between Ω_1^* and $\Omega \times (0, 2\pi/\lambda)$ as follows: if (P, x) is in $\Omega \times (0, 2\pi/\lambda)$ then make correspond to it the point P^* in Ω_1^* defined by $P^* = S_x P$. We define the transformation T on Ω by $TP = S_{2\pi/\lambda}P$. Then this correspondence obviously carries S_t on Ω_1^* into that flow on $\Omega \times (0, 2\pi/\lambda)$ which is built on the transformation T on Ω. As in the proof of Theorem 2 this correspondence gives rise to a measure, $\bar{m}(\bar{M})$, on Ω M $(0, 2\pi/\lambda)$ and the only thing remaining to be shown is that this measure is the desired sort of product measure. To prove this it is sufficient, by Theorem 1, to show that the functions $F(\bar{P})$ and $G(\bar{P})$ defined by (2.1) and (2.2) are $\bar{m}$-measurable. Now $G(\bar{P})$ is $\bar{m}$-measurable because the set in Ω_1^* corresponding to the $\bar{P}$-set for which $G(\bar{P}) < b$ is the P^*-set for which $\arg \psi(P^*) < b\lambda$. $F(\bar{P})$ is $\bar{m}$-measurable because it is a constant.

---

[37] See X, p. 625.

[38] See IV, Lemma 9.1, p. 27.

[39] This remark is due to Koopman, V, p. 318.

To conclude we point out that our first representation theorem can be used to generalize a theorem of von Neumann's about the spectral family E_λ associated with the group U_t. Let $\mathfrak{E}$ be the closed linear manifold spanned by the eigenfunctions of the group U_t. Then E_λ is still a spectral family when considered only on $L_2 - \mathfrak{E}$, the orthogonal complement of $\mathfrak{E}$[40]. The set of points of increase[41] of this spectral family when it is considered only on $L_2 - \mathfrak{E}$ is called the *band spectrum* of the original spectral family.

THEOREM 4.[42] *For a measurable ergodic flow the band spectrum is either the null set or the whole real line.*

PROOF: We shall not give a detailed proof of this theorem. We merely remark that von Neumann proved this assuming the flow to be differentiable in order to find,—in case the flow was not rotation on the circle,—cross sections whose translations over large time intervals were all disjoint. Using our representation theorem (and an unpublished but easily proved lemma of P. R. Halmos to the effect that on a space containing sets of arbitrarily small measure there exists, for every ergodic measure preserving transformation, T, and positive integer, n, a measurable set of positive measure whose first n transforms by T are all disjoint) we are able,—in case the flow is not rotation on the circle,—to obtain such cross sections without any differentiability assumptions. As soon as this is noticed von Neumann's proof can be carried through exactly as in the differentiable case.[43]

THE INSTITUTE FOR ADVANCED STUDY

BIBLIOGRAPHY

I. Birkhoff, G. D.: *Proof of the ergodic theorem*, P. N. A. S., v. 18, (1932), p. 650.
II. Birkhoff, G. D.: *Dynamical Systems*, A. M. A. Colloquium Pub., 1927.
III. Doob, J. L.: *One-parameter families of transformations*, Duke Jnl., v. 4, (1938), p. 752.
IV. Hopf, E.: *Ergodentheorie*, J. Springer, Berlin, 1937.
V. Koopman, B. O.: *Hamiltonian systems and linear transformations in Hilbert space*, P. N. A. S., v. 17, (1931), p. 315.
VI. Saks, S.: *Theory of the Integral*, G. E. Stechert, New York, 1938.
VII. Stone, M. H.: *On one parameter unitary groups in Hilbert space*, Ann. of Math., v. 33, (1932), p. 643.
VIII. Stone, M. H.: *Linear transformations in Hilbert space*, A. M. S. Colloquium Pub., 1932.
IX. von Neumann, J.: *Uber einen satz von Herrn M. H. Stone*, Ann. of Math., v. 33, (1932), p. 567.
X. von Neumann, J.: *Zur operatorenmethode in der klassischen mechanik*, Ann. of Math., v. 33, (1932), p. 587.
XI. Wiener, N.: *The ergodic theorem*, Duke Jnl., v. 5, (1939), p. 1.

[40] See VIII, Theorem 5.13, p. 189.
[41] See VIII, p. 184.
[42] See X, Theorem 5, p. 635.
[43] It has been pointed out to me by Dr. P. R. Halmos that this proof is not correct as it stands, but it has also been pointed out to me by Prof. S. Kakutani that it is possible to make it correct by some minor alterations in the definition of the function $\varphi(P)$.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON THE EXPONENTS OF DIFFERENTIAL IDEALS

BY ELLIS ROBERT KOLCHIN

(Received April 4, 1940)

INTRODUCTION

In the theory of polynomial ideals, in algebra, there are methods, stemming from the theorem of M. Noether, and associated with the names of E. Bertini, E. Lasker, F. S. Macauley, K. Hentzelt, H. Kapferer and P. Dubreil, for finding the exponent of an ideal, or at least a bound for the exponent.

When one seeks to create a notion of exponent for ideals of differential polynomials, one is forced, because of a situation revealed by H. W. Raudenbush,[1] to admit infinite exponents as well as finite ones. The investigation of such exponents, finite or infinite, to some extent for differential ideals of a general character, and to a deeper extent for differential ideals generated by a form in one unknown of the first order, is the object of the present paper.

If we may refer to §2 below for a definition of the *exponent of a differential ideal*, we shall proceed to enumerate our results.

Part II, which presents what is possibly the most interesting portion of our work, deals with differential ideals generated by a single form A, in one unknown, of the first order. The concept of *multiplicity* of a singular solution of A is introduced (§6) and it is shown (§7) that *if A has a singular solution of multiplicity exceeding unity, then the differential ideal generated by A has exponent infinity*. Forms A which have singular solutions, all of multiplicity unity, are discussed in §§8–10. Such singular solutions are divided into two classes, and, guided by a general theorem due to J. F. Ritt, we secure a decomposition of the differential ideal generated by A which puts these two classes into evidence (§8). In §10 it is proved, under an additional assumption (*regular type*) that *the exponent of the differential ideal generated by A is unity or two according as all the singular solutions are in the first class or at least one singular solution is in the second class*. The differential ideal generated by a form A which has *no* singular solutions, is shown, under a certain additional assumption, to have exponent *unity* (§11). Part II concludes with a discussion of a type of form which we call *hyperelliptic*. The "intermediate ideals" are found, and they are shown to fall into *chains* of a fixed length.

Part III contains a brief discussion of chains of differential ideals. A theorem is proved which gives a bound for the exponent of a so-called *principal* chain in terms of the length of the chain (§14).

Part I begins with a statement, in abstract form, of a decomposition theorem due to Ritt. After the definitions of *relative exponent* and other terms, there

[1] Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 371–373.

is proved a theorem connecting the relative exponent of a differential ideal with those of the factor ideals in the decomposition just mentioned (Theorem 1). Also, a result is presented which relates the existence of a strong basis in a differential ideal to the existence of such a basis in the factor ideals (Theorem 2). The remainder of Part I deals with special questions whose treatment is important for the sequel.

In Part I, use is made of the theory of resolvents. This theory, as developed by Ritt, applies to a differential field whose elements are meromorphic functions. To permit the use of abstract fields, there is given, in an appendix, a discussion which supplements Ritt's proofs, making them valid in any differential field of characteristic zero.

Notation

Differentiation (of elements in a differential ring[2]) is indicated by subscripts; thus a_j means the j-th derivative of a. If a single letter is used with different subscripts to denote different ring elements, then differentiation is indicated by a second subscript; thus, the j-th derivative of a_i is a_{ij}.

Square brackets [ ], curled brackets { }, and parentheses ( ), when not used as symbols of aggregation, will mean, respectively, the differential ideal, the perfect differential ideal, and the (ordinary, algebraic) ideal generated by the set of elements they include. These various ideals are supposed to be formed in a fixed differential ring which underlies the discussion. When forming an (algebraic) ideal, the differential ring is considered as an (algebraic) ring.

Membership in a set is denoted in the usual way by the symbol ϵ. Set inclusion is indicated by $\subseteq$, proper inclusion by $\subset$. The symbol for the intersection of sets is $\bigcap$.

The notations

$$a \equiv b \quad (m, n, \cdots),$$
$$a \equiv b \quad [m, n, \cdots],$$
$$a \equiv b \quad \{m, n, \cdots\}$$

will mean, respectively,

$$a - b \,\epsilon\, (m, n, \cdots), \qquad a - b \,\epsilon\, [m, n, \cdots], \qquad a - b \,\epsilon\, \{m, n, \cdots\}.$$

The product of a finite number of differential ideals is defined as their product when considered as (algebraic) ideals. This is readily seen to be a differential ideal. If $\sigma_1, \sigma_2, \cdots, \sigma_s$ are differential ideals, their product is denoted by $\sigma_1\sigma_2 \cdots \sigma_s$.

If σ is a *perfect* differential ideal in a differential ring $\mathfrak{D}$, and if $d \,\epsilon\, \mathfrak{D}$, then

---

[2] The definitions of differential ring, differential ideal, and other terms, are to be found in a paper by Raudenbush, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 361–368.

$\sigma:d$, the quotient of σ by d, is defined as in algebra to be the totality of elements $a \in \mathfrak{D}$ such that $ad \in \sigma$. $\sigma:d$ is itself a perfect differential ideal.

If $\mathfrak{F}$ is a differential field, then $\mathfrak{F}\{u, \cdots\}$ means the differential ring obtained by the differential ring adjunction of $u, \cdots$. The result of differential field adjunction will be denoted by $\mathfrak{F}\langle u, \cdots \rangle$.

As is customary, the word *form* will be used as an abbreviation for *differential polynomial.*

By a *solution* of a set of forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$, is meant a set of elements $\eta_1, \cdots, \eta_n$ of some differential extension field of $\mathfrak{F}$, which annul the forms of the set when substituted for $y_1, \cdots, y_n$, respectively.

Part I. Some General Theorems

1. The product representation

Consider a differential domain of integrity $\mathfrak{D}$ which contains the rational numbers. We suppose that every infinite system in $\mathfrak{D}$ has a basis. The set of forms in a fixed finite number of unknowns, with coefficients in an underlying differential field of characteristic zero, is such a differential domain of integrity.[3]

A set of differential ideals in $\mathfrak{D}$ will be called *separated* if $1 \equiv 0 \ (\sigma_1, \sigma_2)$ for each pair of distinct ideals σ_1, σ_2 in the set.

A differential ideal will be said to be *connected* if it is not the intersection of two separated differential ideals. Since the two ideals are required to be separated, "intersection" may be replaced by "product."[4]

A theorem due to Ritt may be formulated abstractly as follows:

Let σ be a differential ideal in $\mathfrak{D}$. Then σ is connected if and only if $\{\sigma\}$ is. If σ is not connected, it has a representation as the intersection (= product) of a separated finite set of connected differential ideals. This representation is unique. If $\sigma_1\sigma_2 \cdots \sigma_s$ is the representation of σ, then $\{\sigma_1\}\{\sigma_2\} \cdots \{\sigma_s\}$ is the representation for $\{\sigma\}$.[5]

We shall call the representation described in this theorem the *product representation* of σ. Each σ_i will be called a *factor* of the product representation.

2. Exponents

Lemma 1: *Let σ and τ be differential ideals in $\mathfrak{D}$ such that $\sigma \subseteq \tau \subseteq \{\sigma\}$. Then the product representation of σ has the same number of factors as that of τ, and, with a suitable assignment of subscripts, we may write for these representations*

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_s, \qquad \tau = \tau_1\tau_2 \cdots \tau_s, \tag{1}$$
$$\sigma_i \subseteq \tau_i \subseteq \{\sigma_i\}, \qquad i = 1, \cdots, s.$$

Proof: The theorem of §1 shows that we may write for the product representations

---

[3] See the paper cited under footnote 2.

[4] See B. L. van der Waerden, *Moderne Algebra*, vol. 2, Berlin, 1931, p. 46.

[5] Ritt, Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 90–91.

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_s, \qquad \tau = \tau_1\tau_2 \cdots \tau_s,$$
$$\{\sigma_i\} = \{\tau_i\}, \qquad i = 1, \cdots, s.$$

Here product is the same as intersection, and $\sigma \subseteq \tau \subseteq \tau_1$, so that

$$\sigma = (\sigma_1 \cap \tau_1) \cap \sigma_2 \cap \cdots \cap \sigma_s = (\sigma_1 \cap \tau_1)\sigma_2 \cdots \sigma_s.^{6}$$

Since the product representation is unique, this implies that $\sigma_1 = \sigma_1 \cap \tau_1$, that is $\sigma_1 \subseteq \tau_1$. Similarly, every $\sigma_i \subseteq \tau_i$.

Consider two differential ideals σ and τ in $\mathfrak{D}$. If there is a positive integer q such that $\tau^q \subseteq \sigma$, let p be the least such q; if no such q exists, let $p = \infty$. We shall call p the *exponent of σ with respect to τ*, and shall denote it by $l_\tau\sigma$. The exponent of σ with respect to $\{\sigma\}$ will be called, simply, the *exponent of σ*.

In discussing $l_\tau\sigma$ we limit ourselves to cases for which $\sigma \subseteq \tau \subseteq \{\sigma\}$. If $\tau \not\subseteq \{\sigma\}$, then $l_\tau\sigma = \infty$; if $\sigma \not\subseteq \tau$, we may let $\tau' = (\sigma, \tau)$, and then $\sigma \subseteq \tau'$, $l_\tau\sigma = l_{\tau'}\sigma$.

THEOREM 1: *Let σ and τ be differential ideals in $\mathfrak{D}$ such that $\sigma \subseteq \tau \subseteq \{\sigma\}$. Let the product representations of σ and τ be given by* (1). *Then*

$$l_\tau\sigma = \max_{i=1,\cdots,s} l_{\tau_i}\sigma_i.$$

PROOF: Let $p = \max l_{\tau_i}\sigma_i$, $q = l_\tau\sigma$. If $p = \infty$, then $q \leqq p$. Suppose $p < \infty$. Then

$$\tau^p = (\tau_1 \cdots \tau_s)^p = \tau_1^p \cdots \tau_s^p \subseteq \sigma_1 \cdots \sigma_s = \sigma,$$

so that in either case $q \leqq p$. It remains to prove that $p \leqq q$.

If $q = \infty$, then $p \leqq q$. Suppose $q < \infty$. Then

$$\sigma_1^q \cdots \sigma_s^q = (\sigma_1 \cdots \sigma_s)^q = \sigma^q \subseteq \tau = \tau_1 \cdots \tau_s.$$

But $\sigma_1^q, \cdots, \sigma_s^q$ are obviously the factors of the product representation of $\sigma_1^q \cdots \sigma_s^q$. Hence, by Lemma 1, $\sigma_i^q \subseteq \tau_i$, $i = 1, \cdots, s$, that is, $p \leqq q$.

3. Strong bases

In this section we present a theorem which will not be used in the rest of the paper, but which may be of some interest on its own account.

Let σ be a differential ideal in $\mathfrak{D}$. If σ has a strong[7] basis $b_1, \cdots, b_k$, let $q = l_\sigma[b_1, \cdots, b_k]$; of all integers so obtained, let p be the minimum. If σ has no strong basis, let $p = \infty$. We shall call p the *basis index* of σ.

---

[6] It is easy to see that $\sigma_1 \cap \tau_1, \sigma_2, \cdots, \sigma_s$ are separated. For example, because $\sigma_1, \sigma_2, \cdots, \sigma_s$ are separated, there are elements $h_1 \in \sigma_1$ and $h_2 \in \sigma_2$ such that $1 = h_1 + h_2$. Choosing, as we may, p large enough so that $h_1^p \in \tau_1$, we have $h_1^p \in \sigma_1 \cap \tau_1$, $ph_1^{p-1}h_2 + \cdots + h_2^p \in \sigma_2$ and $1 = h_1^p + ph_1^{p-1}h_2 + \cdots + h_2^p$.

[7] A basis of σ is said to be *strong* if there is a single integer such that g^t is in the differential ideal generated by the basis, for every $g \in \sigma$. There exist differential ideals which do not have a strong basis. See Theorem 5 below; also, Kolchin, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 923–926.

THEOREM 2: *Let the differential ideal σ in $\mathfrak{D}$ have the product representation $\sigma = \sigma_1 \cdots \sigma_s$. Then the basis index of σ equals the maximum of the basis indices of the σ_i.*

PROOF: We know that there exist elements m_i, with derivative $m_{i1} \in \sigma$, such that $\sigma_i = (\sigma, m_i)$, $i = 1, \cdots, s$.[8] Let $b_1, \cdots, b_h$ be a strong basis of σ, with $l_\sigma[b_1, \cdots, b_h] = p$. Then

$$\sigma_i^p = (\sigma, m_i)^p \subseteq (\sigma^p, m_i) \subseteq [b_1, \cdots, b_h, m_i].$$

Thus, for each i, $b_1, \cdots, b_h, m_i$ is a strong basis for σ_i, and $l_{\sigma_i}[b_1, \cdots, b_h, m_i] \leqq p$. Hence the basis index of σ is not less than that of each σ_i. Now let $b_1^{(i)}, \cdots, b_{h_i}^{(i)}$ be a strong basis of σ_i, and let $q \geqq l_{\sigma_i}[b_1^{(i)}, \cdots, b_{h_i}^{(i)}]$, $i = 1, \cdots, s$. Since $\sigma_i = (\sigma, m_i)$, we may write, for each i and j,

$$b_j^{(i)} = t_j^{(i)} + c_j^{(i)} m_i, \qquad t_j^{(i)} \in \sigma.$$

Then

$$\begin{aligned}\sigma^q = (\sigma_1 \cdots \sigma_s)^q &= \sigma_1^q \cdots \sigma_s^q \\ &\subseteq [b_1^{(1)}, \cdots, b_{h_1}^{(1)}] \cdots [b_1^{(s)}, \cdots, b_{h_s}^{(s)}] \\ &\subseteq [t_1^{(1)}, t_2^{(1)}, \cdots, t_{h_s}^{(s)}, m_{11}, \cdots, m_{s1}, m_1 \cdots m_s].\end{aligned}$$

Thus, σ has a strong basis whose associated exponent is less than or equal to q, so that the basis index of σ is not more than the greatest basis index of the σ_i.

4. Invariance of the exponent under differential field adjunction

In this section we specialize $\mathfrak{D}$ to a differential ring of forms.

Let $\mathfrak{F}$ be a differential field of characteristic zero. Let $\mathfrak{F}'$ be a differential extension field of $\mathfrak{F}$: $\mathfrak{F} \subseteq \mathfrak{F}'$. Throughout the present section y_i, u_i, w will denote unknowns.

If Φ is a set of forms with coefficients in $\mathfrak{F}$, we shall use (Φ), $[\Phi]$, $\{\Phi\}$ to denote the algebraic, the differential, and the perfect differential ideals, respectively, generated by Φ in the differential ring of forms with coefficients in $\mathfrak{F}$. If Φ is a set of forms with coefficients in $\mathfrak{F}'$, the algebraic, the differential, and the perfect differential ideals generated by Φ in the differential ring of forms with coefficients in $\mathfrak{F}'$ will be indicated, respectively, by $(\Phi)'$, $[\Phi]'$, and $\{\Phi\}'$. It is clear that if Σ is a differential ideal of forms with coefficients in $\mathfrak{F}$, then $(\Sigma)' = [\Sigma]'$.

THEOREM 3: *Let Σ and Λ be differential ideals in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then*

$$l_\Lambda \Sigma = l_{(\Lambda)'}(\Sigma)'.$$

PROOF: Let $p = l_\Lambda \Sigma$, $q = l_{(\Lambda)'}(\Sigma)'$. If $q = \infty$, then $p \leqq q$. Suppose $q < \infty$, and let $G \in \Lambda$. Then $G \in (\Lambda)'$, so that $G^q \in (\Sigma)'$. But $G^q \in \mathfrak{F}\{y_1, \cdots, y_n\}$. Hence $G^q \in \Sigma$,[9] so that $p \leqq q$.

---

[8] Ritt, loc. cit. under footnote 5.

[9] See proof in van der Waerden, loc. cit., p. 67.

If $p = \infty$, then $q \leqq p$. Suppose $p < \infty$, and let $G \in (\Lambda)'$. Then there is a relation

$$G = \sum_{\mu=1}^{m} H_\mu G_\mu, \qquad G_\mu \in \Lambda.$$

Hence

$$G^p = \sum_{i_1+\cdots+i_m=p} \frac{p!}{i_1! \cdots i_m!} H_1^{i_1} \cdots H_m^{i_m} G_1^{i_1} \cdots G_m^{i_m} \in (\Lambda^p)' \subseteq (\Sigma)',$$

so that $q \leqq p$. This completes the proof.

Theorem 3, by itself, yields no information as to how the exponent of a differential ideal is affected by a differential field adjunction, for there is no obvious reason why $\{\Sigma\}'$ should equal $(\{\Sigma\})'$. These two ideals, however, are identical. We devote the rest of this section to the proof of this fact. We use three lemmas.

LEMMA 2: *Let Ω be a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$, with basic set*

$$A, A_1, \cdots, A_p \tag{2}$$

introducing in succession $w, y_1, \cdots, y_p$. Let A be algebraically irreducible over $\mathfrak{F}$, and let each A_i be of order zero and degree unity in y_i. Then (2) is a basic set for $\{\Omega\}'$.

PROOF: Let $B_1, \cdots, B_s$ be the irreducible factors of A over $\mathfrak{F}'$. Let S and I be the separant and initial, respectively, of A; let I_i be the initial (= separant) of A_i. Each B_i is of the same order (call it r) in w as A is, for otherwise A would be reducible over $\mathfrak{F}$. Let T_i be the separant, J_i the initial, of B_i. We have

$$I = J_1 \cdots J_s,$$

$$S = T_1 B_2 \cdots B_s + B_1 T_2 \cdots B_s + \cdots + B_1 B_2 \cdots T_s.$$

Now, $\Omega = \{A, A_1, \cdots, A_p\} : ISI_1 \cdots I_p$. Let

$$\Omega_i = \{B_i, A_1, \cdots, A_p\} : J_i T_i I_1 \cdots I_p, \qquad i = 1, \cdots, s.$$

Then each Ω_i is a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$.[10] Moreover,

---

[10] PROOF: Let $CD \in \Omega_i$. For appropriate k we may write $(I_1 \cdots I_p)^k C \equiv C'$, $(I_1 \cdots I_p)^k D \equiv D' [A_1, \cdots, A_p]$ where C' and D' are forms free of $y_1, \cdots, y_p$. Of course $C'D' \in \Omega_i$. We prove one of C', D' is divisible by B_i. Suppose neither is. Then there are forms M and N, free of $y_1, \cdots, y_p$, such that $E = M(C'D'T_iJ_iI_1 \cdots I_p) + NB_i$, where E is a nonzero form free of $y_1, \cdots, y_p$, of order less than r in w. Clearly $E \in \{B_i, A_1, \cdots, A_p\}$. It readily follows that $(T_iI_1 \cdots I_p)^l E \in (B_i)$ for some positive integer l. Since B_i is irreducible, this implies that $E \in (B_i)$. But the order of E in w is less than r. Hence $E = 0$. This contradiction shows that either C' or D' is divisible by B_i, say C' is. Then $C \in \Omega_i$. Thus, Ω_i is prime.

$$\begin{aligned}
\{\Omega\}' &= \{\{A, A_1, \cdots, A_p\}:ISI_1 \cdots I_p\}' \\
&= \{A, A_1, \cdots, A_p\}':ISI_1 \cdots I_p \\
&= (\{B_1, A_1, \cdots, A_p\}' \cap \cdots \cap \{B_s, A_1, \cdots, A_p\}'):ISI_1 \cdots I_p \\
&= (\{B_1, A_1, \cdots, A_p\}':ISI_1 \cdots I_p) \cap \cdots \\
&\qquad \cap (\{B_s, A_1, \cdots, A_p\}':ISI_1 \cdots I_p) \\
&= (\{B_1, A_1, \cdots, A_p\}':J_1 \cdots J_sT_1B_2 \cdots B_sI_1 \cdots I_p) \\
&\qquad \cap \cdots \cap (\{B_s, A_1, \cdots, A_p\}':J_1 \cdots J_sB_1B_2 \cdots T_sI_1 \cdots I_p) \\
&= (\Omega_1:J_2 \cdots J_sB_2 \cdots B_s) \cap \cdots \cap (\Omega_s:J_1 \cdots J_{s-1}B_1 \cdots B_{s-1}) \\
&= \Omega_1 \cap \cdots \cap \Omega_s.^{11}
\end{aligned}$$

$\{\Omega\}'$ does not contain a form of class $q+1$ of order less than r in w, because Ω_i does not. Let C be of class $q+1$ and have order r in w, and suppose that $C \in \{\Omega\}'$. $C \in \Omega_i$, and is therefore divisible by B_i, $i = 1, \cdots, s$. Hence C is divisible by A. From this it follows that (2) is a basic set for $\{\Omega\}'$.

LEMMA 3: *Let $\mathfrak{F}$ contain non-constant elements.*[12] *Let Σ be a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with arbitrary unknowns $u_1, \cdots, u_q$.*[13] *Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: It suffices to show that $\{\Sigma\}' \subseteq (\Sigma)'$. Let $A = 0$ be a resolvent for Σ, and let Ω be the associated prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$, with basic set (2), where A is of order r in w, and each A_i is of order zero and degree unity in y_i.[14] We know that

$$\Sigma = \Omega \cap \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}.$$

Let $G \in \{\Sigma\}'$. We may write

$$G = G_0 + G_1\omega_1 + \cdots + G_m\omega_m$$

where each $G_\mu \in \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, and where $1, \omega_1, \cdots, \omega_m$ are elements of $\mathfrak{F}'$, linearly independent over $\mathfrak{F}$. We shall show that each $G_\mu \in \Sigma$.

Let S and I be the separant and initial of A; let I_i be the initial of A_i. Evidently there exist non-negative integers a, b, a_j such that

$$I^aS^bI_1^{a_1} \cdots I_p^{a_p}G_\mu \equiv R_\mu(\Omega), \qquad \mu = 0, 1, \cdots, m, \tag{3}$$

where each R_μ is reduced with respect to the basic set (2). Then

$$I^aS^bI_1^{a_1} \cdots I_p^{a_p}G \equiv R_0 + R_1\omega_1 + \cdots + R_m\omega_m(\Omega)'. \tag{4}$$

But, by Lemma 2, (2) is a basic set for $\{\Omega\}'$. Hence the second member of (4),

[11] The parentheses in these equations are symbols of aggregation.

[12] An element of $\mathfrak{F}$ is called *constant*, if its derivative vanishes.

[13] The u_i need not actually occur.

[14] See Appendix.

which is in $\{\Omega\}'$, yet reduced with respect to (2), must vanish. Since $1, \omega_1, \cdots, \omega_m$ are linearly independent over $\mathfrak{F}$, it follows that each R_μ vanishes. (3) then shows that each $G_\mu \in \Omega$. Since G_μ is also in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, we have $G_\mu \in \Sigma$, $\mu = 0, 1, \cdots, m$.

Hence $G \in (\Sigma)'$, so that $\{\Sigma\}' \subseteq (\Sigma)'$.

LEMMA 4: *Let $\mathfrak{F}$ consist purely of constants; let $\mathfrak{F}' = \mathfrak{F}\langle x\rangle$, where x is an element whose derivative is 1. Let Σ be a prime differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: We must show that $\{\Sigma\}' \subseteq (\Sigma)'$. Let $G \in \{\Sigma\}'$. For a suitable non-zero element $\varphi \in \mathfrak{F}\langle x\rangle$, we may write

$$\varphi G = H_0 + H_1 x + \cdots + H_m x^m,$$

where the H_μ are in $\mathfrak{F}\{y_1, \cdots, y_n\}$. For a large integer r, $G^r \in (\Sigma)'$. Hence

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} \in (\Sigma)'.$$

Thus, we see that we may write

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} = d_1 S_1 + \cdots + d_k S_k,$$

where each $S_i \in \Sigma$, and each $d_i \in \mathfrak{F}\langle x\rangle$. Let $\omega_1, \cdots, \omega_t$ be elements of $\mathfrak{F}\langle x\rangle$ such that $1, x, \cdots, x^{mr}, \omega_1, \cdots, \omega_t$ are linearly independent over $\mathfrak{F}$, and such that each d_i is linearly dependent on $1, x, \cdots, x^{mr}, \omega_1, \cdots, \omega_t$. Then there is a relation

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} = T_0 + \cdots + T_{mr} x^{mr} + U_1\omega_1 + \cdots + U_t\omega_t,$$

where the T_i and U_i are in Σ. By the linear independence it follows that $H_0^r = T_0$, so that $H_0 \in \Sigma$. Hence

$$H_1 + H_2 x + \cdots + H_m x^{m-1} \in \{\Sigma\}'.$$

Continuing, we find every $H_\mu \in \Sigma$. Therefore $G \in (\Sigma)'$ and $\{\Sigma\}' \subseteq (\Sigma)'$.

We are now in a position to prove our theorem. $\mathfrak{F}$ is again any differential field of characteristic zero, and $\mathfrak{F}'$ a differential extension thereof.

THEOREM 4: *Let Σ be a perfect differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: Let $\Sigma = \Sigma_1 \cap \cdots \cap \Sigma_s$ be the decomposition of Σ into essential prime differential ideals. It is easy to see that

$$\{\Sigma\}' = \{\Sigma_1\}' \cap \cdots \cap \{\Sigma_s\}'. \tag{5}$$

We shall prove that

$$(\Sigma)' = (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'. \tag{6}$$

Indeed, since $\Sigma \subseteq \Sigma_i$, we have $(\Sigma)' \subseteq (\Sigma_i)'$, $i = 1, \cdots, s$, that is

$$(\Sigma)' \subseteq (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'.$$

Now let $G \in (\Sigma_i)'$, $i = 1, \cdots, s$. We may write

$$\begin{aligned}
\{\Omega\}' &= \{\{A, A_1, \cdots, A_p\}:ISI_1 \cdots I_p\}' \\
&= \{A, A_1, \cdots, A_p\}':ISI_1 \cdots I_p \\
&= (\{B_1, A_1, \cdots, A_p\}' \cap \cdots \cap \{B_s, A_1, \cdots, A_p\}'):ISI_1 \cdots I_p \\
&= (\{B_1, A_1, \cdots, A_p\}':ISI_1 \cdots I_p) \cap \cdots \\
&\qquad \cap (\{B_s, A_1, \cdots, A_p\}':ISI_1 \cdots I_p) \\
&= (\{B_1, A_1, \cdots, A_p\}':J_1 \cdots J_sT_1B_2 \cdots B_sI_1 \cdots I_p) \\
&\qquad \cap \cdots \cap (\{B_s, A_1, \cdots, A_p\}':J_1 \cdots J_sB_1B_2 \cdots T_sI_1 \cdots I_p) \\
&= (\Omega_1:J_2 \cdots J_sB_2 \cdots B_s) \cap \cdots \cap (\Omega_s:J_1 \cdots J_{s-1}B_1 \cdots B_{s-1}) \\
&= \Omega_1 \cap \cdots \cap \Omega_s.^{11}
\end{aligned}$$

$\{\Omega\}'$ does not contain a form of class $q+1$ of order less than r in w, because Ω_i does not. Let C be of class $q+1$ and have order r in w, and suppose that $C \in \{\Omega\}'$. $C \in \Omega_i$, and is therefore divisible by B_i, $i = 1, \cdots, s$. Hence C is divisible by A. From this it follows that (2) is a basic set for $\{\Omega\}'$.

LEMMA 3: *Let $\mathfrak{F}$ contain non-constant elements.*[12] *Let Σ be a prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with arbitrary unknowns $u_1, \cdots, u_q$.*[13] *Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: It suffices to show that $\{\Sigma\}' \subseteq (\Sigma)'$. Let $A = 0$ be a resolvent for Σ, and let Ω be the associated prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}$, with basic set (2), where A is of order r in w, and each A_i is of order zero and degree unity in y_i.[14] We know that

$$\Sigma = \Omega \cap \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}.$$

Let $G \in \{\Sigma\}'$. We may write

$$G = G_0 + G_1\omega_1 + \cdots + G_m\omega_m$$

where each $G_\mu \in \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, and where $1, \omega_1, \cdots, \omega_m$ are elements of $\mathfrak{F}'$, linearly independent over $\mathfrak{F}$. We shall show that each $G_\mu \in \Sigma$.

Let S and I be the separant and initial of A; let I_i be the initial of A_i. Evidently there exist non-negative integers a, b, a_j such that

$$I^aS^bI_1^{a_1} \cdots I_p^{a_p}G_\mu \equiv R_\mu(\Omega), \qquad \mu = 0, 1, \cdots, m, \tag{3}$$

where each R_μ is reduced with respect to the basic set (2). Then

$$I^aS^bI_1^{a_1} \cdots I_p^{a_p}G \equiv R_0 + R_1\omega_1 + \cdots + R_m\omega_m(\Omega)'. \tag{4}$$

But, by Lemma 2, (2) is a basic set for $\{\Omega\}'$. Hence the second member of (4),

---

[11] The parentheses in these equations are symbols of aggregation.

[12] An element of $\mathfrak{F}$ is called *constant*, if its derivative vanishes.

[13] The u_i need not actually occur.

[14] See Appendix.

which is in $\{\Omega\}'$, yet reduced with respect to (2), must vanish. Since $1, \omega_1, \cdots, \omega_m$ are linearly independent over $\mathfrak{F}$, it follows that each R_μ vanishes. (3) then shows that each $G_\mu \in \Omega$. Since G_μ is also in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, we have $G_\mu \in \Sigma$, $\mu = 0, 1, \cdots, m$.

Hence $G \in (\Sigma)'$, so that $\{\Sigma\}' \subseteq (\Sigma)'$.

LEMMA 4: *Let $\mathfrak{F}$ consist purely of constants; let $\mathfrak{F}' = \mathfrak{F}\langle x\rangle$, where x is an element whose derivative is 1. Let Σ be a prime differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: We must show that $\{\Sigma\}' \subseteq (\Sigma)'$. Let $G \in \{\Sigma\}'$. For a suitable non-zero element $\varphi \in \mathfrak{F}\langle x\rangle$, we may write

$$\varphi G = H_0 + H_1 x + \cdots + H_m x^m,$$

where the H_μ are in $\mathfrak{F}\{y_1, \cdots, y_n\}$. For a large integer r, $G^r \in (\Sigma)'$. Hence

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} \in (\Sigma)'.$$

Thus, we see that we may write

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} = d_1 S_1 + \cdots + d_k S_k,$$

where each $S_i \in \Sigma$, and each $d_i \in \mathfrak{F}\langle x\rangle$. Let $\omega_1, \cdots, \omega_t$ be elements of $\mathfrak{F}\langle x\rangle$ such that $1, x, \cdots, x^{mr}, \omega_1, \cdots, \omega_t$ are linearly independent over $\mathfrak{F}$, and such that each d_i is linearly dependent on $1, x, \cdots, x^{mr}, \omega_1, \cdots, \omega_t$. Then there is a relation

$$H_0^r + rH_0^{r-1}H_1 x + \cdots + H_m^r x^{mr} = T_0 + \cdots + T_{mr}x^{mr} + U_1\omega_1 + \cdots + U_t\omega_t,$$

where the T_i and U_i are in Σ. By the linear independence it follows that $H_0^r = T_0$, so that $H_0 \in \Sigma$. Hence

$$H_1 + H_2 x + \cdots + H_m x^{m-1} \in \{\Sigma\}'.$$

Continuing, we find every $H_\mu \in \Sigma$. Therefore $G \in (\Sigma)'$ and $\{\Sigma\}' \subseteq (\Sigma)'$.

We are now in a position to prove our theorem. $\mathfrak{F}$ is again any differential field of characteristic zero, and $\mathfrak{F}'$ a differential extension thereof.

THEOREM 4: *Let Σ be a perfect differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then $(\Sigma)' = \{\Sigma\}'$.*

PROOF: Let $\Sigma = \Sigma_1 \cap \cdots \cap \Sigma_s$ be the decomposition of Σ into essential prime differential ideals. It is easy to see that

$$\{\Sigma\}' = \{\Sigma_1\}' \cap \cdots \cap \{\Sigma_s\}'. \tag{5}$$

We shall prove that

$$(\Sigma)' = (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'. \tag{6}$$

Indeed, since $\Sigma \subseteq \Sigma_i$, we have $(\Sigma)' \subseteq (\Sigma_i)'$, $i = 1, \cdots, s$, that is

$$(\Sigma)' \subseteq (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'.$$

Now let $G \in (\Sigma_i)'$, $i = 1, \cdots, s$. We may write

$$G = G_0^{(i)} + G_1^{(i)}\omega_1 + \cdots + G_m^{(i)}\omega_m,$$

where each $G_i^{(j)} \in \Sigma_j$, and $1, \omega_1, \cdots, \omega_m$ are elements of $\mathfrak{F}'$, linearly independent over $\mathfrak{F}$. The linear independence shows that $G_i^{(1)} = \cdots = G_i^{(s)}$, $i = 0, 1, \cdots, m$, so that each $G_i^{(j)} \in \Sigma_1 \cap \cdots \cap \Sigma_s = \Sigma$, and $G \in (\Sigma)'$. Hence

$$(\Sigma)' \supseteq (\Sigma_1)' \cap \cdots \cap (\Sigma_s)'.$$

This proves (6).

If $\mathfrak{F}$ contains non-constants, then, by (5), (6) and Lemma 3, $\{\Sigma\}' = (\Sigma)'$.

Suppose $\mathfrak{F}$ consists purely of constants. Let $\mathfrak{F}^* = \mathfrak{F}\langle x\rangle$, where x is an element whose derivative is 1, and let $\mathfrak{F}^\ddagger = \mathfrak{F}'\langle x\rangle$. By Lemma 4 and equations analogous to (5) and (6),

$$\begin{aligned}(\Sigma)^* &= (\Sigma_1)^* \cap \cdots \cap (\Sigma_s)^* \\ &= \{\Sigma_1\}^* \cap \cdots \cap \{\Sigma_s\}^* = \{\Sigma\}^*.\end{aligned}$$

Hence, by the part of the theorem already proved,

$$\begin{aligned}(\Sigma)^\ddagger &= ((\Sigma)^*)^\ddagger = (\{\Sigma\}^*)^\ddagger \\ &= \{\{\Sigma\}^*\}^\ddagger = \{\Sigma\}^\ddagger.\end{aligned}$$

Thus, $(\Sigma)^\ddagger$ is a perfect differential ideal.

Now, $(\Sigma)^\ddagger = ((\Sigma)')^\ddagger$. Suppose that $(\Sigma)' \neq \{\Sigma\}'$. Then there is an $F \in \mathfrak{F}'\{y_1, \cdots, y_n\}$ such that $F \in \{\Sigma\}'$, $F \notin (\Sigma)'$. Clearly $F \in \{\Sigma\}^\ddagger$, that is, $F \in (\Sigma)^\ddagger = ((\Sigma)')^\ddagger$. Since $F \in \mathfrak{F}'\{y_1, \cdots, y_n\}$, this implies that $F \in (\Sigma)'$. This contradiction shows that $(\Sigma)' = \{\Sigma\}'$, and completes the proof.

COROLLARY 1: *Let Σ be a differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then the exponent of Σ equals that of $(\Sigma)'$.*

PROOF: By Theorems 3 and 4

$$l_{\{\Sigma\}}\Sigma = l_{(\{\Sigma\})'}(\Sigma)' = l_{\{\{\Sigma\}\}'}(\Sigma)' = l_{\{\Sigma\}'}(\Sigma)'.$$

5. A certain differential ideal

Let $\mathfrak{F}$ be any differential field of characteristic zero, and let y be an unknown. We work in the differential ring $\mathfrak{F}\{y\}$. The following theorem uses an idea due to Raudenbush.[15]

THEOREM 5: *$[y]^2$ does not have a strong basis.*

PROOF: Assume that $[y]^2$ has a strong basis. Then it has a strong basis consisting purely of forms $y_i y_j$. Let s be an integer so large that the set of forms

$$y_i y_j, \qquad 0 \leqq i \leqq j \leqq s, \tag{7}$$

is a strong basis of $[y]^2$. Let p be the exponent with respect to $[y]^2$ of the differential ideal generated by the basis (7). Let α be a positive integer whose value (large) is to be given later.

[15] See paper cited under footnote 1.

Consider the forms

$$y_{i_1}y_{i_2}\cdots y_{i_{2p}}, \qquad i_1+\cdots+i_{2p}=\alpha. \tag{8}$$

Each form (8) is in $[y]^{2p}$. Considering degree and weight, we see that each form (8) is a linear combination, with coefficients in $\mathfrak{F}$, of forms

$$(y_iy_j)_k y_{i_1}\cdots y_{i_{2p-2}},$$

$$0\leqq i\leqq j\leqq s, \qquad i+j+k+j_1+\cdots+j_{2p-2}=\alpha. \tag{9}$$

Hence the number of linearly independent forms (8) does not now exceed the number of forms (9). Since the totality of forms (8) are linearly independent, it follows that the number of distinct forms (8) is less than or equal to the number of forms (9).

Now, the number of expressions (8) is clearly at least

$$\frac{1}{(2p)!}\left(\frac{\alpha}{2p}+1\right)^{2p-1}$$

if we suppose that α is divisible by $2p$. For, we may assign $i_1, \cdots, i_{2p-1}$ arbitrary values from 0 to $\alpha/2p$, and then i_{2p} is uniquely determined; moreover, the number of times a given form (8) can be produced in this way is at most $(2p)!$ On the other hand, the number of expressions (9) is surely not more than

$$\tfrac{1}{2}(s+1)(s+2)(\alpha+1)^{2p-2},$$

for the number of forms in the basis (7) is $(s+1)(s+2)/2$. But if α is sufficiently large, then

$$\frac{1}{(2p)!}\left(\frac{\alpha}{2p}+1\right)^{2p-1} > \tfrac{1}{2}(s+1)(s+2)(\alpha+1)^{2p-2}.$$

Hence, the number of forms (8) exceeds the number of forms (9). This contradiction completes the proof.

COROLLARY 2: *Let $P\in\mathfrak{F}\{y\}$ be of order zero. Then the exponent of $[P]$ is ∞ or 1 according as P has or has not a multiple factor.*

PROOF: By Corollary 1 we may suppose that $\mathfrak{F}$ contains all the roots of $P=0$. Let

$$P=b(y-\eta_1)^{k_1}\cdots(y-\eta_s)^{k_s}, \qquad \eta_i\neq\eta_j \text{ if } i\neq j,$$

be the factorization of P into linear factors. It is easy to see that

$$[P]=[(y-\eta_1)^{k_1}]\cdots[(y-\eta_s)^{k_s}] \tag{10}$$

is the product representation of $[P]$. If $k_i=1$, then $[(y-\eta_i)^{k_i}]=[y-\eta_i]$ obviously has exponent unity. If $k_i>1$, then $[(y-\eta_i)^{k_i}]$ has exponent ∞. For, otherwise $(y-\eta_i)^{k_i}$ would be a strong basis for $[y-\eta_i]^2$, a contradiction of Theorem 5. Since by Theorem 1 the exponent of $[P]$ equals the maximum of the exponents of the $[(y-\eta_i)^{k_i}]$, the proof is complete.

PART II. THE DIFFERENTIAL IDEAL GENERATED BY A FORM IN ONE UNKNOWN OF ORDER UNITY

6. Preliminary remarks

Consider a form A in a single unknown y, of order unity. We seek the exponent of $[A]$. We shall work in a fixed differential field $\mathfrak{F}$ of characteristic zero, which contains the coefficients of A. By Corollary 1, we may (and do) assume, without further mention, that $\mathfrak{F}$ is extensive enough to contain certain solutions of A which we shall discuss.

The exponent of $[A]$ will be found to depend on the nature of the *singular solutions* of A, that is, the solutions of A which annul S, the separant of A.[16] A singular solution of A must annul the resultant with respect to y_1, the derivative of y, of A and S. This resultant is a nonzero form of order zero, or is 0, according as A has no multiple factor of order unity, or has such a factor, that is, according as A and S have not or have a common factor of order unity. In the former case, A has only a finite number of singular solutions, and these are easily found.

If $y = \eta$ is a singular solution of A, we consider A as a form in $z = y - \eta$, and write

$$A = \sum_{\nu=0}^{n} P_\nu z^{p_\nu} z_1^\nu, \tag{11}$$

where each P_ν either is of order zero and not divisible by z, or vanishes, and where the p_ν are definite non-negative integers.[17] We suppose that the degree of A in y_1 is n, so that $P_n \neq 0$.

Let m be the minimum of the total degrees of all terms effectively present in A considered as a polynomial in z and z_1. We shall call m the *multiplicity* of the singular solution $y = \eta$. It is obvious that if A has a multiple factor, any solution of that factor is a singular solution of multiplicity greater than unity.

7. Singular solutions of multiplicity exceeding unity

We shall prove that *if A has a singular solution of multiplicity exceeding unity, then the exponent of $[A]$ is ∞.*

(a) Let $y = \eta$ be a singular solution of A of multiplicity at least 2, and suppose that $y = \eta$ is an essential manifold of A. Let $z = y - \eta$. The product representation of $[A]$ must have a factor whose manifold is $y = \eta$, that is, a factor Σ_1 such that $\{\Sigma_1\} = \{z\}$. Now, it is known that

$$\Sigma_1 = [A, z^q], \tag{12}$$

[16] The definition here of *singular solution* is broader than that given by Ritt in his book, *Differential equations from the algebraic standpoint*, and used in the Appendix of the present paper, for A is not assumed here to be algebraically irreducible.

[17] If $P_\nu = 0$, we let p_ν be any fixed non-negative integer.

where q is any sufficiently large integer.[18] Because $y = \eta$ is of multiplicity at least 2, it follows that $A \in [z^2, z_1^2]$. Hence, by (12),

$$\Sigma_1 \subseteq [z^2, z_1^2].$$

Thus, the exponent of Σ_1 is not less than that of $[z^2, z_1^2]$, which, by Theorem 5, is ∞. Therefore, using Theorem 1, we see that the exponent of $[A]$ is ∞.

(b) Suppose that $y = \eta$ is a singular solution of A of multiplicity greater than unity, which is not an essential manifold of A. Let $y - \eta = z$. We shall show first that $\{A\}$ contains a form which, when considered as a form in z, has at least one term of the first degree.

Let $C_1, \cdots, C_q$ be the distinct irreducible factors of A of order unity, and let $F = C_1 \cdots C_q$. If T denotes the separant of F, the derivative of F is a form $Tz_2 + U$, where U is a form of order not exceeding unity. Hence

$$Tz_2 \equiv -U \qquad [F].$$

It follows, by successive differentiations and eliminations, that

$$T^{a_i} z_i \equiv V_i \qquad [F], \qquad i = 2, \cdots, n+2,$$

where the a_i are appropriate integers, and the V_i are forms whose order is not greater than unity. Letting a be the maximum of the a_i, we have

$$T^a z_i \equiv W_i \qquad [F], \qquad i = 2, \cdots, n+2,$$

where $W_i = T^{a-a_i} V_i$.

Let J be the initial of F. Then J is of order zero. Clearly, there exists a non-negative integer b such that

$$J^b W_i \equiv X_i \qquad (F), \qquad i = 2, \cdots, n+2,$$

where each X_i is of order not exceeding unity, and is of degree less than n in z_1.[19] It follows that

$$J^b T^a z_i \equiv X_i \qquad [F], \qquad i = 2, \cdots, n+2.$$

The $n+1$ forms X_i are linear combinations, with coefficients which are polynomials in z, of the n quantities $1, z_1, \cdots, z_1^{n-1}$. Hence, some linear combination of the X_i, with coefficients which are forms of order zero, not all 0, must vanish:

$$\sum_{i=2}^{n+2} K_i X_i = 0.$$

We assume, as we may, that the K_i are not all divisible by z. Referring to the last congruence, we now see that

---

[18] Ritt and Kolchin, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 895–898.

[19] The degree of F in z_1 is at most n.

$$J^b T^a M \equiv 0 \qquad [F],$$

where $M = \sum_{i=2}^{n+2} K_i z_i$.

Since not every K_i is divisible by z, M has at least one term of degree unity. Consider the form $y - \zeta$, where ζ is either a common solution of TJ and F, or a root of a factor of A of order zero, but is distinct from η. Let N be the product of all such forms $y - \zeta$, and let $Z = MN$.

Z, like M, has at least one term of degree unity (when considered as a form in z). Z vanishes for all solutions of F with the possible exception of $y = \eta$, and for all the solutions of A which are not solutions of F, again with the possible exception of $y = \eta$. That is, Z vanishes for all solutions of A, save possibly $y = \eta$. Since $y = \eta$ is not an essential manifold, Z must also vanish for $y = \eta$. By the analog to the Hilbert-Netto theorem,[20] then, $Z \in \{A\}$. Z is the form whose existence we were to prove.

Thus, $\{A\}$ contains a form $z_j + H$, where H, considered as a form in z, has each of its terms either of order less than j or of degree greater than unity.

Assume, now, that the exponent of $[A]$ is finite, say q. Then[21]

$$(z_{j+i_1} + H_{i_1}) \cdots (z_{j+i_q} + H_{i_q}) \equiv 0 \qquad [A]$$

for all non-negative integers $i_1, \cdots, i_q$. Since $y = \eta$ is of multiplicity at least 2, as a singular solution of A, it follows that $A \in [z^2, z_1^2]$. Hence

$$(z_{j+i_1} + H_{i_1}) \cdots (z_{j+i_q} + H_{i_q}) \equiv 0_i \qquad [z^2, z_1^2].$$

If we compare terms here of degree q and weight $qj + i_1 + \cdots + i_q$, we obtain

$$z_{j+i_1} \cdots z_{j+i_q} \equiv 0 \qquad [z^2, z_1^2], \qquad i_1, \cdots, i_q = 0, 1, 2, \cdots.$$

This is easily seen to imply, however, that $[z]^2$ has a strong basis. This contradiction of Theorem 5 completes the proof.

EXAMPLE 1. Let $A = \sum_{\nu=0}^{n} Q_\nu y_1^\nu$ $(n \geqq 1)$, where the Q_ν are forms of order zero with *constant* coefficients, and where $Q_n \neq 0$.

If A has no multiple factors of order unity, the singular solutions must be constants, for they must satisfy an algebraic equation with constant coefficients;[22] since $S = \sum_{\nu=1}^{n} \nu Q_\nu y_1^{\nu-1}$, it is evident that the singular solutions of A are precisely the common zeros of Q_0 and Q_1. A singular solution is of multiplicity exceeding unity if it is a multiple zero of Q_0. Thus, *if A has a multiple factor, or if Q_0 has a multiple factor which is also a factor of Q_1, then the exponent of $[A]$ is ∞.*

8. Singular solutions of multiplicity unity; the product representation

Throughout §§8–10 we suppose that A has singular solutions, and that they are all of multiplicity unity. In the present section we shall show that all the

[20] See the paper cited under footnote 2.

[21] H_{i_k} is the i_k-th derivative of H.

[22] The equation obtained by eliminating z_1 from $A = 0$, $S = 0$.

singular solutions are essential manifolds, and shall find the product representation of $[A]$. With an additional assumption, we shall determine the exponent of $[A]$ (§10).

Since A has no singular solution of multiplicity exceeding unity, A has no multiple factors. Hence A has only a finite number of singular solutions (§6).

Let $y = \eta$ be a singular solution of A, and write $z = y - \eta$. Since S must vanish for $y = \eta$, we see from (11) that either $P_1 = 0$ or $p_1 > 0$. This means, since $y = \eta$ is of multiplicity unity, that $P_0 \not\equiv 0$ and $p_0 = 1$. (11) thus becomes

$$A = P_0 z + \sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^\nu. \tag{13}$$

It now follows from a theorem of Ritt that $y = \eta$ is an essential manifold of A.[23]

We now turn to the product representation of $[A]$.

Let $y = \eta_i$, $i = 1, \cdots, s$, be the singular solutions of A; we denote $y - \eta_i$ by u_i. It is evident that we may write

$$[A] = \Sigma_0 \Sigma_1 \cdots \Sigma_s, \tag{14}$$

where the manifold of Σ_0 consists of the non-singular solutions of A,[24] and where the manifold of Σ_i, $i > 0$, is $y = \eta_i$, that is, $\{\Sigma_i\} = \{u_i\}$, $i = 1, \cdots, s$.

Referring to (13) and making use of a known result,[25] we see that

$$\Sigma_i = [u_i], \qquad i = 1, \cdots, s. \tag{15}$$

The determination of Σ_0 is more difficult. The first step will be to find $\{\Sigma_0\}$. If z is one of the u_i, then (13) holds. Differentiating (13) we obtain

$$A_1 = P_{01} z + \sum_{\nu=1}^{n} P_{\nu 1} z^{p_\nu} z_1^\nu + z_1 \left(P_0 + \sum_{\nu=1}^{n} p_\nu P_\nu z^{p_\nu - 1} z_1^\nu \right) + \sum_{\nu=1}^{n} \nu P_\nu z^{p_\nu} z_1^{\nu-1} \cdot z_2,$$

where $P_{\nu 1}$ is the derivative of P_ν. But by (13),

$$P_0 z \equiv -z_1 \sum_{\nu=1}^{n} P_\nu z^{p_\nu} z_1^{\nu-1} \qquad (A). \tag{16}$$

Hence

$$P_0 A_1 = P_{01} P_0 z + P_0 z_1 \sum_{\nu=1}^{n} P_{\nu 1} z^{p_\nu} z_1^{\nu-1} + P_0 z_1 \left(P_0 + \sum_{\nu=1}^{n} p_\nu P_\nu z^{p_\nu - 1} z_1^\nu \right)$$
$$+ \left(P_1 z^{p_1 - 1} \cdot P_0 z + P_0 z_1 \sum_{\nu=2}^{n} \nu P_\nu z^{p_\nu} z_1^{\nu-2} \right) z_2$$

---

[23] Ritt, Annals of Mathematics, vol. 37 (1936), pp. 552–617. See especially §5. Professor Ritt's proof is not applicable to abstract differential fields, as it involves function-theoretic concepts. But an abstract proof of the *sufficiency* part of the theorem (we need only the sufficiency) has recently been achieved by H. Levi, and may be expected to appear in the literature shortly.

[24] If A is algebraically irreducible, the manifold of Σ_0 is the general solution of A.

[25] See §2 of the paper cited under footnote 18.

$$\equiv z_1\Big(-P_{01}\sum_{\nu=1}^{n}P_\nu z^{p_\nu}z_1^{\nu-1}+P_0\sum_{\nu=1}^{n}P_{\nu 1}z^{p_\nu}z_1^{\nu-1}+P_0\Big(P_0+\sum_{\nu=1}^{n}p_\nu P_\nu z^{p_\nu-1}z_1^{\nu}\Big)$$

$$+\Big(-P_1z^{p_1-1}\sum_{\nu=1}^{n}P_\nu z^{p_\nu}z_1^{\nu-1}+P_0\sum_{\nu=2}^{n}\nu P_\nu z^{p_\nu}z_1^{\nu-2}\Big)z_2\Big), \qquad (A),$$

so that

$$P_0A_1 \equiv z_1(Uz_2+V) \qquad (A), \tag{17}$$

where

$$U=-P_1z^{p_1-1}\sum_{\nu=1}^{n}P_\nu z^{p_\nu}z_1^{\nu-1}+P_0\sum_{\nu=2}^{n}\nu P_\nu z^{p_\nu}z_1^{\nu-2},$$

$$V=-P_{01}\sum_{\nu=1}^{n}P_\nu z^{p_\nu}z_1^{\nu-1}+P_0\sum_{\nu=1}^{n}P_{\nu 1}z^{p_\nu}z_1^{\nu-1}+P_0\Big(P_0+\sum_{\nu=1}^{n}p_\nu P_\nu z^{p_\nu-1}z_1^{\nu}\Big). \tag{18}$$

From (17) and (16) we find that

$$z(P_0Uz_2+P_0V)\equiv 0 \qquad (A, A_1).$$

This implies that

$$\{A\}=\{z\}\cap\{A, P_0Uz_2+P_0V\}.$$

The two ideals in the second member here are separated, for by (18)

$$P_0V\equiv P_0^3\not\equiv 0 \qquad [z], \tag{19}$$

so that $P_0Uz_2+P_0V\not\equiv 0$ $[z]$.

Now, z represents any one of the u_i. Let

$$I_i=P_0U, \qquad J_i=P_0V \quad \text{when} \quad z=u_i, \qquad i=1,\cdots,s. \tag{20}$$

Then by the above, for $i=1,\cdots,s$,

$$\{A\}=\{u_i\}\cap\{A, I_iu_{i2}+J_i\},$$

$$u_i(I_iu_{i2}+J_i)\equiv 0, \qquad u_{i1}(I_iu_{i2}+J_i)\equiv 0 \qquad (A, A_1), \tag{21}$$

and $y=\eta_i$ is not a solution of $I_iu_{i2}+J_i$. It readily follows that

$$\{A\}=\{A, I_1u_{12}+J_1,\cdots,I_su_{s2}+J_s\}\cap\{u_1\}\cap\cdots\cap\{u_s\}. \tag{22}$$

The $s+1$ differential ideals, whose intersection appears in the second member of (23), are separated, so that we may replace intersection by product.

Comparing (22) with

$$\{A\}=\{\Sigma_0\}\cap\{\Sigma_1\}\cap\cdots\cap\{\Sigma_s\}$$

(which follows from (14)) and with (15), we see that

$$\{\Sigma_0\}=\{A, I_1u_{12}+J_1,\cdots,I_su_{s2}+J_s\}. \tag{23}$$

We shall now obtain two forms M and N such that

(24) $M \epsilon \{u_1\} \cdots \{u_s\}, \quad N \epsilon \{\Sigma_0\}, \quad MN \epsilon [A], \quad M + N = 1.$

This will enable us to determine Σ_0.

Let

$$B_i = I_i u_{i2} + J_i, \qquad i = 1, \cdots, s,$$

and let

$$Q_i = P_0 \quad \text{when} \quad z = u_i, \qquad i = 1, \cdots, s.$$

Then Q_i is not divisible by u_i, and, by (19) and (20),

(25) $$B_i - Q_i^3 \equiv I_i u_{i2} \quad (u_i, u_{i1}).$$

Since Q_i is not divisible by u_i, there is a form C_i of order zero, and an element $a_i \epsilon \mathfrak{F}$, such that

$$1 = a_i Q_i^3 + C_i u_i .^{26}$$

This may be written

$$1 = [C_i u_i - a_i(B_i - Q_i^3)] + a_i B_i .$$

By (21), (25) and the definition of B_i, moreover,

$$[C_i u_i - a_i(B_i - Q_i^3)] \cdot a_i B_i \equiv 0 \qquad [A],$$

provided $I_i \equiv 0 \ (u_i, u_{i1})$, that is, provided $P_2 = 0$ or $p_2 > 0$ when $z = u_i$ (we refer to (20) and (18), and the fact that $p_1 > 0$). If this is not the case, then

$$[C_i u_i - a_i(B_i - Q_i^3)] \cdot (a_i B_i)^2 \equiv 0 \qquad [A],$$

because $u_{i2} B_i^2 \equiv 0$ is a consequence of $u_{i1} B_i \equiv 0$, and the latter congruence holds by (21).

Accordingly, we separate the singular solutions of A into two classes. A singular solution $y = \eta$ of A will be said to be of the *first class* if in A, considered as a form in $z = y - \eta$, the coefficient of z_1^2 is divisible by z; a singular solution which is not of the first class will be said to be of the *second class.*

We suppose the singular solutions η_i ordered so that $\eta_1, \cdots, \eta_r$ are of the first class, and $\eta_{r+1}, \cdots, \eta_s$ are of the second class. This is the same as supposing that

(26)
$$\begin{aligned} &P_2 = 0 \quad \text{or} \quad p_2 > 0 \quad \text{when} \quad z = u_i, && i = 1, \cdots, r, \\ &P_2 \neq 0 \quad \text{and} \quad p_2 = 0 \quad \text{when} \quad z = u_i, && i = r+1, \cdots, s. \end{aligned}$$

We let

(27)
$$\begin{aligned} M_i &= C_i u_i - a_i(B_i - Q_i^3), && i = 1, \cdots, r, \\ M_i &= [C_i u_i - a_i(B_i - Q_i^3)]^2 \\ &\quad + 2[C_i u_i - a_i(B_i - Q_i^3)] \cdot a_i B_i, && i = r+1, \cdots, s, \\ N_i &= a_i B_i, && i = 1, \cdots, r, \\ N_i &= (a_i B_i)^2, && i = r+1, \cdots, s. \end{aligned}$$

[26] It will be observed that $a_i \neq 0$.

Then $M_i \in \{u_i\}$, $N_i \in \{\Sigma_0\}$, $M_i + N_i = 1$, $M_iN_i \in [A]$, $i = 1, \cdots, s$. Letting, therefore,

$$(28) \qquad \begin{aligned} M &= M_1 \cdots M_s \\ N &= (M_1 + N_1) \cdots (M_s + N_s) - M_1 \cdots M_s = 1 - M, \end{aligned}$$

we see that (24) holds.

We refer now to the paper cited under footnote 5. The proof in that paper shows that $\Sigma_0 = [A, N]$. But, by (28), $N \equiv 0$ $(N_i, \cdots, N_s)$, and $N_i = MN_i + NN_i \equiv 0$ $[A, N]$. Hence $\Sigma_0 = [A, N_1, \cdots, N_s]$. Referring to (27), we see, finally, that

$$(29) \qquad \Sigma_0 = [A, B_1, \cdots, B_r, B_{r+1}^2, \cdots, B_s^2].$$

The product representation of A is given by (14), (15) *and* (29).[27]

9. Continuation

The purpose of this section is to prove that

$$(30) \qquad [A, B_1, \cdots, B_s]^2 \subseteq \Sigma_0.$$

By (29) it suffices to show that

$$(31) \qquad B_{\sigma i}B_{\tau j} \equiv 0 \ (\Sigma_0), \qquad \sigma, \tau = 1, \cdots, s, \qquad i, j = 0, 1, 2, \cdots.$$

Let $Iz_2 + J$ represent any $B_\sigma = I_\sigma u_{\sigma 2} + J_\sigma$. We shall prove that

$$(32) \qquad (Iz_2 + J)_i(Iz_2 + J)_j \equiv 0 \qquad (\Sigma_0)$$

for all i, j.

From (21) we see that

$$(33) \qquad z(Iz_2 + J) \equiv 0, \qquad z_1(Iz_2 + J) \equiv 0 \qquad [A].$$

Differentiating the second congruence here $l + 1$ times, we find that

$$(34) \qquad \sum_{\lambda=0}^{l+1} \binom{l+1}{\lambda} z_{\lambda+1}(Iz_2 + J)_{l+1-\lambda} \equiv 0 \qquad [A].$$

By (29), (32) holds for $i = 0, j = 0$. We assume that (32) holds for $i = 0, 1, \cdots, l - 1$ and $j = 0$, and prove it for $i = l, j = 0$.

Multiplying (34) by $Iz_2 + J$, we obtain, using (33) and the induction assumption,

$$(l + 1)z_2(Iz_2 + J)_l(Iz_2 + J) \equiv 0 \qquad (\Sigma_0).$$

But $Iz_2(Iz_2 + J) \equiv -J(Iz_2 + J)$, (Σ_0). Hence

$$J(Iz_2 + J)_2(Iz_2 + J) \equiv 0 \qquad (\Sigma_0).$$

---

[27] Actually, Σ_0 may not be connected. This will occur if A has more than one factor of order unity.

However, by (20) and (19), J is congruent, modulo (z, z_1), to a nonzero element of $\mathfrak{F}$. Hence, by (33), $(Iz_2 + J)_l(Iz_2 + J) \equiv 0\ (\Sigma_0)$.

Thus (32) holds for all i when $j = 0$.

To prove (32) for all i and all j we again use induction. Assume (32) to be valid for all i and for $j = l - 1$. Differentiating (32) for $j = l - 1$ we see that

$$(Iz_2 + J)_{i+1}(Iz_2 + J)_{l-1} + (Iz_2 + J)_i(Iz_2 + J)_l \equiv 0 \qquad (\Sigma_0),$$

so that, by the induction assumption, $(Iz_2 + J)_i(Iz_2 + J)_l \equiv 0\ (\Sigma_0)$. This proves (32) for all i and j.

Thus (31) holds whenever $\sigma = \tau$. To prove (31) in the general case, consider $B_{\sigma i}B_{\tau j}$. It is sufficient, of course, to suppose that $\sigma > r$, for $B_\sigma \in \Sigma_0$ if $\sigma \leqq r$. Now, the remainder of $B_{\tau j}$ with respect to B_σ is a form in $\{\Sigma_0\}$ of order not exceeding unity. Such a form must be divisible by A. Hence, for an appropriate integer m, $I_\sigma^m B_{\tau j} \equiv 0\ [A, B_\sigma]$. Hence, since (31) holds when $\tau = \sigma$, $I_\sigma^m B_{\sigma i}B_{\tau j} \equiv 0\ (\Sigma_0)$. But $\sigma > \tau$, so that (see (26), (20) and (18)) I_σ is congruent modulo $(u_\sigma, u_{\sigma 1})$, to a nonzero element of $\mathfrak{F}$. Hence, by (21), $B_{\sigma i}B_{\tau j} \equiv 0\ (\Sigma_0)$.

10. The regular type

We shall say that A is of the *regular type* if

$$1 \equiv 0\ (\Sigma_0, S, I_1, \cdots, I_s). \tag{35}$$

In the present section *we assume that A is of the regular type.*

We shall prove that *the exponent of $[A]$ is 1 or 2 according as all of the singular solutions of A are of the first class or at least one singular solution of A is of the second class.*

(a) Let all the singular solutions of A be of the first class, and let $G \in \{\Sigma_0\}$. We shall show that $G \in \Sigma_0$.

For each i we have, for an appropriate integer b_i, $I_i^{b_i}G \equiv G_i'[B_i]$, where G_i' is of order unity, at most. Since $G_i' \in \{\Sigma_0\}$, G_i' must be divisible by A. Hence, for $i = 1, \cdots, s$, $I^{b_i}G \equiv 0\ [A, B_i]$. Similarly, for a suitable c, $S^cG \equiv 0\ [A]$. Now, by (35), there are forms X_i, and a form $T \in \Sigma_0$, such that

$$1 = T + X_0S^c + X_1I_1^{b_1} + \cdots + X_sI_s^{b_s}.$$

Therefore,

$$G = (T + X_0S^c + X_1I_1^{b_1} + \cdots + X_sI_s^{b_s})G \equiv 0 \qquad [A, B_1, \cdots, B_s].$$

But all the singular solutions of A are of the first class. Therefore, by (29), $\Sigma_0 = [A, B_1, \cdots, B_s]$. Hence, $G \in \Sigma_0$.

Thus, the exponent of Σ_0 is unity. But the exponent of $\Sigma_i = [u_i]$, $i = 1, \cdots, s$, is obviously unity. Hence, by Theorem 1, unity is also the exponent of $[A]$.

(b) Let A have at least one singular solution of the second class, and let $G \in \{\Sigma_0\}$. As in (a) we see that $G \in [A, B_1, \cdots, B_s]$. This, together with (31), proves that the exponent of Σ_0 does not exceed 2. Hence the exponent of $[A]$ is no greater than 2.

To prove that the exponent of $[A]$ is exactly 2, it suffices to exhibit a form in $\{A\}$ which is not in $[A]$. To this end, we let z be any u_i with $i > r$, and show that $\{\Sigma_0\}$ contains a form $z_h + H$, where h is a sufficiently large positive integer, and H is a form, of order less than h, contained in $[z]$.

By (35) there is a relation

$$1 = T + X_0S + X_1I_1 + \cdots + X_sI_s, \qquad T \in \Sigma_0. \tag{36}$$

Let $h(\geqq 3)$ be an integer such that $h - 1$ exceeds the orders of $T, X_0, \cdots, X_s$. If we differentiate each B_i, written as a form in z, $h - 3$ times, we obtain relations $I_iz_{h-1} + K_i \equiv 0 \ \{\Sigma_0\}$. Similarly, differentiating A $h - 2$ times we obtain $Sz_{h-1} + K_0 \equiv 0 \ \{\Sigma_0\}$. All K_i are forms of order less than $h - 1$. These relations, together with $Tz_{h-1} \equiv 0\{\Sigma_0\}$, yield, because of (36), a relation $z_{h-1} + K \equiv 0 \ \{\Sigma_0\}$, where K is a form of order less than $h - 1$.

Let α be the term of K free of z and the derivatives of z. If $\alpha = 0$, then $(z_{h-1} + K)_1$ is a form $z_h + H$ as described above; if $\alpha \neq 0$, then $\alpha(\alpha^{-1}(z_{h-1} + K))_1$ is such a form.

Now let E be the product of all the u_i other than z, and consider $E(z_h + H)$, where $z_h + H$ is the form whose existence we have just proved. We see that $\{A\}$ contains a form $Ez_h + F$, where E is a form of order zero, not divisible by z, and F is a form of order less than h.

We shall now complete the proof by showing that $Ez_h + F \not\equiv 0\,[A]$. The method will be to assume that $Ez_h + F \equiv 0\,[A]$, and to force a contradiction.

Let then

$$Ez_h + F = C_0A + C_1A_1 + \cdots + C_tA_t, \tag{37}$$

where A_j represents the j-th derivative of A. By (17), for all $i > 0$,

$$P_0A_i \equiv (z_1(Uz_2 + V))_{i-1} \qquad (A, A_1, \cdots, A_{i-1}).$$

It follows from (37), therefore, that

$$\begin{aligned} P_0^t(Ez_h + F) = {} & D_0A + D_1z_1(Uz_2 + V) \\ & + D_2(z_1(Uz_2 + V)_1 + z_2(Uz_2 + V)) \\ & + \cdots + D_t\sum_{i=0}^{t-1}\binom{t-1}{i}z_{i+1}(Uz_2 + V)_{t-1-i}. \end{aligned} \tag{38}$$

We now consider the various forms in (38) as polynomials in the z_i. We shall modify (38) in such a way as to make evident a contradiction

Consider a power series[28] $\sum_{i=1}^{\infty}\beta_iz_1^i$ in z_1 such that A vanishes when $\Sigma\beta_iz_1^i$ is substituted for z. Such a power series exists, as by (13) $z_1 = 0, z = 0$ renders A zero and leaves $\dfrac{\partial A}{\partial z}$ nonzero.[29]

[28] All power series which we mention are formal.

[29] The "implicit function theorem" for abstract fields of characteristic zero may be proved by the method of undetermined coefficients. The proof is simpler than the one in analysis, because the series is formal and no questions of convergence arise.

Substituting $\Sigma\beta_i z_1^i$ for z in (38), we find a relation

$$E'z_h + F' = D_1'z_1(U'z_2 + W_0) + D_2'(z_1(U'z_3 + W_1) + z_2(U'z_2 + W_0)) + \cdots + D_t'\sum_{i=0}^{t-1}\binom{t-1}{i}z_{i+1}(U'z_{t+1-i} + W_{t-1-i}). \tag{39}$$

Here E', F', U', the D_i' and the W_i are power series in z_1 with coefficients which are polynomials in the z_i with $i > 1$. Each W_i is free of z_j with $j > i + 1$. Moreover, E' and U' involve only z_1, and have nonzero terms of degree zero. For E' this is true because $P_0^i E$ is not divisible by z; for U' because $P_2 \neq 0$ and $p_2 = 0$ (z is a u_i with $i > r$; see (26) and (18).)

Let

$$w_1 = z_1, \qquad w_i = U'z_i + W_{i-2}, \qquad i = 2, 3, \cdots. \tag{40}$$

Since U' has a nonzero term of degree zero, the equations (40) may be inverted to obtain

$$z_1 = w_1, \qquad z_i = Yw_i + Z_i, \qquad i = 2, 3, \cdots, \tag{41}$$

where Y is a series in w_1 with a nonzero term of degree zero, and each Z_i is a series in w_1 with coefficients which are polynomials in $w_2, \cdots, w_{i-1}$. The equations (41) define a substitution, with inverse substitution (40). Applying substitution (41) to (39), we obtain a relation

$$Qw_h + R = L_1w_1w_2 + L_2(w_1w_3 + (Yw_2 + Z_2)w_2) + \cdots + L_t(w_1w_{t+1} + (t-1)(Yw_2 + Z_2)w_t + \cdots + (Yw_t + Z_t)w_2), \tag{42}$$

where Q is a series in w_1 with a nonzero term a of degree zero, and R is free of the w_i with $i \geqq h$.

For the second member of (42) to produce the term aw_h which appears in the first member, some L_i with $i \geqq h$ must have a nonzero term of degree zero, because each term in the coefficient of L_i is of at least the first degree in $w_1, w_2, \cdots, w_i$. Thus, some of the L_i have a term which is simply an element of $\mathfrak{F}$, and hence certainly terms bw_1^j ($b \in \mathfrak{F}$, $b \neq 0$, $j \geqq 0$). Let L_p be the L_i of highest subscript, with this property. Then $p \geqq h$. Of all the terms of L_p of the above mentioned type, let cw_1^k be the one of least degree. Then

$$L_p(w_1w_{p+1} + \cdots + (Yw_p + Z_p)w_2)$$

contains the term $cw_1^{k+1}w_{p+1}$. This term does not appear in the first member of (42), and must therefore be cancelled by terms in some expression

$$L_i(w_1w_{i+1} + \cdots + (Yw_i + Z_i)w_2) \tag{43}$$

with $i \neq p$. But if $i < p$, then each term of (43) is divisible by some w_j with $2 \leqq j \leqq p$. Hence $cw_1^{k+1}w_{p+1}$ must be cancelled by terms from expressions (43) with $i > p$. This implies that some L_i with $i > p$ has a term of the type bw_1^{k+1} ($b \in \mathfrak{F}$, $b \neq 0$). This contradicts the definition of L_p, and completes the proof.

The following example shows that of the forms A with *constant coefficients*, which have singular solutions, all of multiplicity unity, those which are not of the regular type are exceptional. It also shows that, for any preassigned pair of integers r, s $(0 \leqq r \leqq s,\ s > 0)$, there exist forms of the regular type having r singular solutions of the first class and $s - r$ singular solutions of the second class.

EXAMPLE 2. Let r, s, m, n be integers with $0 \leqq r \leqq s \leqq m$, $s > 0$, $n \geqq 3$, let $u_i = y - \eta_i$, $i = 1, \cdots, s$, where $\eta_1, \cdots, \eta_s$ are distinct constant elements of a differential field $\mathfrak{E}$ of characteristic zero, and let

$$Q_0 = \sum_{\mu=0}^{m-s} a_{\mu 0} y^{\mu}, \qquad Q_1 = \sum_{\mu=0}^{m-s} a_{\mu 1} y^{\mu}, \qquad Q_2 = \sum_{\mu=0}^{m-r} a_{\mu 2} y^{\mu},$$

$$Q_\nu = \sum_{\mu=0}^{m} a_{\mu\nu} y^{\mu}, \qquad \nu = 3, \cdots, n,$$

where $a_{00}, a_{01}, \cdots, a_{mn}$ is a set of constants, algebraically independent with respect to $\mathfrak{E}$. Consider the form

$$A = Q_0 u_1 \cdots u_s + Q_1 u_1 \cdots u_s y_1 + Q_2 u_1 \cdots u_r y_1^2 + \sum_{\nu=3}^{n} Q_\nu y_1^\nu$$

with coefficients in $\mathfrak{F} = \mathfrak{E}\langle a_{00}, a_{01}, \cdots, a_{mn}\rangle$.

We shall show that the singular solutions of A are $\eta_1, \cdots, \eta_s$, where $\eta_1, \cdots, \eta_r$ are of the first class, and $\eta_{r+1}, \cdots, \eta_s$ are of the second. We shall prove, moreover, that A is of the regular type. In fact, we shall prove the stronger result that

$$1 \equiv 0 \quad (A, S, I_1, \cdots, I_s, J_1, \cdots, J_s). \tag{44}$$

Let R be the resultant with respect to y_1 of A and S. R is a polynomial in y and the $a_{\mu\nu}$. If we let the $a_{\mu\nu}$ take on special values, such that A becomes $u_1 \cdots u_s + y_1^n$, then S becomes $n y_1^{n-1}$, so that R becomes $n^n (u_1 \cdots u_s)^{n-1} \neq 0$. Since R is not 0 for special values of the $a_{\mu\nu}$, R does not vanish for unknown $a_{\mu\nu}$.

Now, each singular solution is a solution of R, which is of order zero and has constant coefficients. Since $R \neq 0$, this implies that all the singular solutions of A are constants. Hence, from the expression for A and the expression for S derived therefrom, we see that a singular solution of A must annul $Q_0 u_1 \cdots u_s$ and $Q_1 u_1 \cdots u_s$, that is, must be one of the η_i. That each η_i is a singular solution is obvious. Thus, the singular solutions of A are $\eta_1, \cdots, \eta_s$. It is apparent that these are of multiplicity unity, and that η_i is of the first or second class according as $i \leqq r$ or $i > r$.

We find, using (20) and (18) that

$$I_i \equiv Q_0 (u_1 \cdots u_{i-1} u_{i+1} \cdots u_s)^2 I, \qquad J_i \equiv (Q_0 u_1 \cdots u_{i-1} u_{i+1} \cdots u_s)^2 J, \quad (A),$$

where

$$I = -Q_1^2 u_1 \cdots u_s - Q_1 Q_2 u_1 \cdots u_r y_1 - Q_1 \sum_{\nu=3}^{n} Q_\nu y_1^{\nu-1}$$

$$+ 2Q_0 Q_2 u_1 \cdots u_r + Q_0 \sum_{\nu=3}^{n} \nu Q_\nu y_1^{\nu-2},$$

$$J = (Q_0 u_1 \cdots u_s)' + (Q_1 u_1 \cdots u_s)' y_1 + (Q_2 u_1 \cdots u_r)' y_1^2 + \sum_{\nu=3}^{n} Q_\nu' y_1^\nu,$$

accents indicating differentiation with respect to y. Hence,

$$(A, S, I_1, \cdots, I_s, J_1, \cdots, J_s) = (A, S, Q_0 I, Q_0^2 J).$$

Suppose, now, that (44) is false. Then there is a pair of values $y = \bar{y}, y_1 = \bar{y}_1$ which annul the polynomials A, S, $Q_0 I$, $Q_0^2 J$. When $y_1 = 0$, A, S and $Q_0^2 J$ become, respectively, $Q_0 u_1 \cdots u_s$, $Q_1 u_1 \cdots u_s$ and $Q_0^2 (Q_0 u_1 \cdots u_s)'$, and these polynomials have no solution in common. Hence $\bar{y}_1 \neq 0$. Also, the resultant with respect to y_1 of S and

$$Q_1 u_1 \cdots u_s + Q_2 u_1 \cdots u_r y_1 + \sum_{\nu=3}^{n} Q_\nu y_1^{\nu-1} = \frac{A - Q_0 u_1 \cdots u_s}{y_1}$$

is a nonzero polynomial in y, relatively prime to Q_0.[30] Hence $\bar{y}$ is not a root of $Q_0 = 0$, so that $\bar{y}, \bar{y}_1$ annul I and J.

Let M be the resultant with respect to y_1, N the resultant with respect to y, of J and

$$nA - y_1 S = nQ_0 u_1 \cdots u_s + (n-1) Q_1 u_1 \cdots u_s y_1$$

$$+ (n-2) Q_2 u_1 \cdots u_r y_1^2 + \sum_{\nu=3}^{n-1} (n-\nu) Q_\nu y_1^\nu.$$

M is a polynomial in y, N is a polynomial in y_1, each with coefficients which are polynomials in the $a_{\mu\nu}$ *not including* a_{0n}. Moreover, M and N are not 0.[31] Now, $\bar{y}, \bar{y}_1$ annul M and N. Thus, $\bar{y}$ and $\bar{y}_1$ are algebraic functions of the $a_{\mu\nu}$ other than a_{0n}. Since A vanishes for $y = \bar{y}, y_1 = \bar{y}_1$, and since $\bar{y}_1 \neq 0$, it follows that a_{0n} is an algebraic function of the other $a_{\mu\nu}$. This contradiction completes the proof.

---

[30] This resultant is nonzero because it is nonzero when $A = y_1^n + u_1 \cdots u_s y_1$; it is prime to Q_0 because its coefficients are independent of $a_{00}, \cdots, a_{m-s,0}$.

[31] If we specialize some of the $a_{\mu\nu}$ so that $A = y_1^n + y_1^{n-1} + Q_0 u_1 \cdots u_s$, then $J = (Q_0 u_1 \cdots u_s)'$, $nA - y_1 S = nQ_0 u_1 \cdots u_s + y_1^n$, and $M = (Q_0 u_1 \cdots u_s)'^{n-1} \neq 0$, so that M does not vanish before we specialize. If we specialize A further by setting $y_1 = 0$, then $J = (Q_0 u_1 \cdots u_s)'$, $nA - y_1 S = nQ_0 u_1 \cdots u_s$, and N is the resultant of $nQ_0 u_1 \cdots u_s$ and $(Q_0 u_1 \cdots u_s)^1$, which is not 0. Thus, N is not 0 before the specialization.

EXAMPLE 3. Let $n \geqq 2$, and let $A = Qy_1^n + P$, where P and Q are relatively prime nonzero forms of order zero with constant coefficients, P being of positive degree and without multiple factors. The singular solutions of A are the roots of $P = 0$, and are all of multiplicity unity.

If $z = y - \eta$, where η is a root of $P = 0$, then $\eta_1 = 0$, and we may write $A = (P/z)z + Qz_1^n$. Thus, referring to (18), we see that

$$U = n(P/z)Qz_1^{n-2},$$
$$V = -(P/z)'Qz_1^n + (P/z)Q'z_1^n + (P/z)^2.$$

Hence, if $\eta_1, \cdots, \eta_s$ are the roots of $P = 0$, then (see (20))

$$I_i = n\left(\frac{P}{y - \eta_i}\right)^2 Qy_1^{n-2},$$
$$J_i = \frac{P}{y - \eta_i}\left(-\left(\frac{P}{y - \eta_i}\right)' Qy_1^n + \frac{P}{y - \eta_i} Q'y_1^n\right) + \left(\frac{P}{y - \eta_i}\right)^3.$$

Since the roots of $P = 0$ are distinct, it follows that $Qy_1^{n-2} \equiv 0(I_1, \cdots, I_s)$, so that $P \equiv 0(A, I_1, \cdots, I_s)$. Since P and Q are relatively prime, then, $y_1^{n-2} \equiv 0(A, I_1, \cdots, I_s)$, so that

$$\left(\frac{P}{y - \eta_i}\right)^3 \equiv 0 \ (A, I_1, \cdots, I_s, J_1, \cdots, J_s), \qquad i = 1, \cdots, s.$$

From this it follows that $1 \equiv 0 \ (A, I_1, \cdots, I_s, J_1, \cdots, J_s)$, so that A is of the regular type.

The coefficient in A of y_1^2 is divisible by $y - \eta_i$ if and only if $n > 2$ (in which case the coefficient vanishes). Thus, the singular solutions of A are all of the first class or are all of the second class according as $n > 2$ or $n = 2$. Consequently, the exponent of $Qy_1^n + P$ is 1 or 2 according as $n > 2$ or $n = 2$.

EXAMPLE 4. Let $A = xy_1^2 - 2yy_1 + y$, where x is any element whose derivative is 1. Then $S = 2(xy_1 - y)$, so that the singular solutions of A are $y = 0$ and $y = x$. We may write $A = x(y_1 - 1)^2 - 2(y - x)(y_1 - 1) - (y - x)$. Both singular solutions are of multiplicity unity, and are of the second class.

If we let $\eta_1 = 0$, $\eta_2 = x$, we have, referring to (20) and (18), $I_1 = 2(xy_1 - 2y + x)$, $I_2 = -2(xy_1 - 2y)$. Hence $1 = (I_1 + I_2)/2x$, so that A is of the regular type, and the exponent of $[A]$ is 2.

The following example shows that not every form with singular solutions, all of multiplicity unity, is of the regular type.

EXAMPLE 5. Let $A = y_1^4 + y_1^3 + P$, where P is a form of order zero, with constant coefficients, and with distinct zeros $\eta_1, \cdots, \eta_s$. The singular solutions of A are $\eta_1, \cdots, \eta_s$, all of multiplicity unity and of the first class. $S = 4y_1^3 + 3y_1^2$. By (18) and (20),

$$I_i = \left(\frac{P}{y - \eta_i}\right)^2 (4y_1^2 + 3y_1),$$

$$J_i = \frac{P}{y-\eta_i}\left(-\left(\frac{P}{y-\eta_i}\right)' y_1(y_1^3+y_1^2) + \left(\frac{P}{y-\eta_i}\right)^2\right)$$
$$\equiv \frac{P}{y-\eta_i}\left(P\left(\frac{P}{y-\eta_i}\right)' + \left(\frac{P}{y-\eta_i}\right)^2\right) \equiv \left(\frac{P}{y-\eta_i}\right)^2 P', \qquad (A).$$

It follows by (29) that $\Sigma_0 = [A, B] = (A, [B])$, where $B = (4y_1^2 + 3y_1)y_2 + P'$. Hence

$$(\Sigma_0, S, I_1, \cdots, I_s) = (y_1^4 + y_1^3 + P, 4y_1^2 + 3y_1, B, B_1, B_2, \cdots)$$
$$= (256P^2 - 27P, 9y_1 + 64P, B, B_1, B_2, \cdots)$$
$$= (256P - 27, P', 9y_1 + 64P, B_1, B_2, \cdots).$$

Thus A is of the regular type if $256P - 27$ and P' are relatively prime. However, if $256P - 27$ and P' have a factor in common which is relatively prime to P'', then A is not of the regular type. This is easy to see from the congruences

$$B_1 = (4y_1^2 + 3y_1)y_3 + (8y_1 + 3)y_2^2 + P''y_1$$
$$\equiv -3y_2^2 - \tfrac{3}{4}P'' \qquad (256P - 27, 9y_1 + 64P),$$
$$B_2 = (4y_1^2 + 3y_1)y_4 + 3(8y_1 + 3)y_2y_3 + \cdots$$
$$\equiv -9y_2y_3 + H_2 \qquad (256P - 27, 9y_1 + 64P),$$
$$B_k \equiv -3(k+1)y_2y_{k+1} + H_k \qquad (256P - 27, 9y_1 + 64P),$$

where H_k is of order not exceeding k.

11. Case in which no singular solution exists

In this section we suppose that A *has no singular solutions.* By the analog of the Hilbert-Netto theorem this means that

$$1 \equiv 0 \ (S, S_1, \cdots, S_p, A, A_1, \cdots, A_q)$$

for sufficiently large integers p and q.

We shall consider only those forms A for which p may be taken as 0, that is, for which there is a q such that

$$1 \equiv 0 \ (S, A, A_1, \cdots, A_q). \qquad (45)$$

We prove that *if* (45) *holds, than the exponent of* $[A]$ *is unity.*

Indeed, let $G \in \{A\}$. It is required to show that $G \in [A]$. Performing a partial reduction of G with respect to A, we see that there is a positive integer j such that $S^jG \equiv G'[A]$, where G' is of order less than or equal to unity. $G' \in \{A\}$, and hence is divisible by A. Therefore $S^jG \equiv 0$ $[A]$. Of course, $A_iG \equiv 0$ $[A]$, $i = 0, 1, \cdots, q$. By (41), however, $1 \equiv 0$ $(S^j, A, A_1, \cdots, A_q)$. Hence, $G = 1 \cdot G \equiv 0$ $[A]$, q.e.d.

The following example shows that a form A, which has no singular solution, for which (45) fails to hold, is exceptional.

EXAMPLE 6. Let $\mathfrak{F}$ be the differential field obtained by adjoining the unknowns $a_{\mu\nu}$ ($\mu = 0, \cdots, m$; $\nu = 0, \cdots, n$) to some differential field of characteristic zero. We consider in $\mathfrak{F}\{y\}$ the form

$$A = \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} a_{\mu\nu} y^{\mu} y_1^{\nu}.$$

We shall show that A has no singular solutions, and that (45) holds, with $q = 1$.

A_1 is linear in y_2. Let $A_1 = Sy_2 + T$. We must prove that $1 \equiv 0(A, S, T)$. Suppose $1 \not\equiv 0(A, S, T)$. Then there is a solution $y = \bar{y}$, $y_1 = \bar{y}_1$ of the polynomials A, S, T. Now, $\bar{y}$ and $\bar{y}_1$ must annul, respectively, the resultant with respect to y_1 and the resultant with respect to y, of S and T. These resultants are nonzero polynomials in y and y_1, respectively, with coefficients which are polynomials in the $a_{\mu\nu}$ and $a_{\mu\nu 1}$, *not including* a_{00}. It follows, since $y = \bar{y}$, $y_1 = \bar{y}_1$ is a solution of A, that a_{00} is an algebraic function of the other $a_{\mu\nu}$ and the $a_{\mu\nu 1}$. This contradiction completes the proof.

There exist forms A for which (45) is satisfied for some $q > 1$, but for which (45) is not satisfied if $q = 1$. This is shown by

EXAMPLE 7. Let $A = 1 + 2y_1 + (1 - y^3)y_1^2$. Then $S = 2 + 2(1 - y^3)y_1$, and A has no singular solutions. Now, $y_1 + 1 = A - (y_1/2)S$, so that

$$(S, A) = (y^3, y_1 + 1).$$

Also, $A_1 \equiv -3y^2y_1^3 + 2(1 + (1 - y^3)y_1)y_2 \equiv 3y^2(S, A)$. Hence

$$(S, A, A_1) = (y^2, y_1 + 1).$$

Again, $A_2 = -6yy_1^4 - 15y^2y_1^2y_2 + 2(1 - y^3)y_2^2$
$+ 2(1 + (1 - y^3)y_1)y_3 \equiv -6y + 2y_2^2\ (S, A, A_1)$,

so that

$$(S, A, A_1, A_2) = (y^2, y_1 + 1, y_2^2 - 3y).$$

Finally, $A_3 = -6y_1^5 - 54yy_1^3y_2 - 36y^2y_1y_2^2 - 21y^2y_1^2y_3 + 6(1 - y^3)y_2y_3 + 2(1 + (1 - y^3)y_1)y_4 \equiv 6(1 + 9yy_2 - y_2y_3)$, (S, A, A_1, A_2), so that $1 \equiv y_2(y_3 - 9y)$, (S, A, A_1, A_2, A_3), and $1 \equiv y_2^4(y_3 - 9y)^4 \equiv 9y^2(y_3 - 9y)^4 \equiv 0$ (S, A, A_1, A_2, A_3). Thus, (45) holds for $q = 3$, but for no lower value of q.

12. The intermediate ideals for hyperelliptic forms

We have seen in §10 that if A has singular solutions, all of multiplicity unity, at least one of which is of the second class, and if A is of the regular type, then $[A]$ is properly included in $\{A\}$, and has exponent 2. It is natural now to ask what differential ideals there are *between* $[A]$ and $\{A\}$, that is, what differential ideals Λ there are such that $[A] \subseteq \Lambda \subseteq \{A\}$. We shall call such a Λ an *intermediate ideal.* In the present section we answer this question completely for the forms of Example 3 with $n = 2$.

We are dealing then with a form $A = Qy_1^2 + P$, where P and Q have constant

coefficients, are of order zero, and are relatively prime. P is assumed to be of positive degree, relatively prime to $P' = \frac{d}{dy}P$. Such a form A we shall call a *hyperelliptic* form.

Our first task will be to find a suitable basis for Σ_0.

Let $\eta_1, \cdots, \eta_s$ be the singular solutions of A, that is, the roots of $P = 0$. We have seen that

$$\Sigma_0 = [A, (I_1y_2 + J_1)^2, \cdots, (I_sy_2 + J_s)^2], \tag{46}$$[32]

where

$$(y - \eta_i)(I_iy_2 + J_i) \equiv 0, \qquad y_1(I_iy_2 + J_i) \equiv 0 \quad [A], \tag{47}$$

and

$$1 \equiv 0\ (A, I_1, \cdots, I_s). \tag{48}$$

Congruence (48) shows that there is a form K, with constant coefficients and of order less than 2, such that

$$y_2 + K \equiv 0\ (A, I_1y_2 + J_1, \cdots, I_sy_2 + J_s). \tag{49}$$

We shall show that

$$\Sigma_0 = [A, (y_2 + K)^2], \qquad \{\Sigma_0\} = [A, y_2 + K], \tag{50}$$

and that

$$(y - \eta_1) \cdots (y - \eta_s)(y_2 + K) \equiv 0, \qquad y_1(y_2 + K) \equiv 0 \quad [A]. \tag{51}$$

Now, (51) is obvious from (47) and (49). As to (50), it is evident from (46), (49) and the fact that Σ_0 is of exponent 2, that $[A, (y_2 + K)^2] \subseteq \Sigma_0$. To prove the inclusion in the opposite direction, consider any $I_iy_2 + J_i$. We have

$$I_iy_2 + J_i = I_i(y_2 + K) + J_i - I_iK.$$

Hence $J_i - I_iK \in \{\Sigma_0\}$. Since the order of $J_i - I_iK$ does not exceed unity, this form is divisible by A. Thus,

$$I_i y_2 + J_i \equiv 0\ (A, y_2 + K),$$

so that

$$(I_iy_2 + J_i)^2 \equiv 0\ [A, (y_2 + K)^2], \qquad i = 1, \cdots, s.$$

This, in conjunction with (46), shows that $\Sigma_0 \subseteq [A, (y_2 + K)^2]$, and completes the proof of the first part of (50). The second part of (50) is now obvious.

The congruence (50) gives the basis sought for Σ_0. We have, as a result,

---

[32] Since P has constant coefficients the η_i are constants, so that $(y - \eta_i)_j = y_j$, $i = 1, \cdots, s$, $j = 1, 2, \cdots$.

$$(52) \qquad [A] = [P]\cdot[A, (y_2 + K)^2], \qquad \{A\} = [P]\cdot[A, y_2 + K],$$

for P is a constant times the product of the $y - \eta_i$.[33]

If we refer to (51), the first intermediate ideals which present themselves are of the type

$$(53) \qquad [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_g})(y_2 + K)],$$

where $i_1, \cdots, i_g$ are distinct integers from 1 to s. If $g = s$, the ideal (53) is $[A]$, if $g = 0$, (53) is $\{A\}$.

We shall show that *the ideals* (53) *are distinct from one another*, and that *every intermediate ideal is one of the ideals* (53), that is, the ideals (53) are the only intermediate ideals.

To prove that the ideals (53) are all distinct, it suffices to show, $h = 1, \cdots, s$, that

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h-1}})(y_2 + K)] \neq [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)].$$

For, if

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)]$$

and

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{j_1}) \cdots (y - \eta_{j_f})(y_2 + K)]$$

were equal, they would both equal

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{k_1}) \cdots (y - \eta_{k_l})(y_2 + K)],$$

where $k_1, \cdots, k_l$ are the integers common to $i_1, \cdots, i_h$ and $j_1, \cdots, j_f$; and if

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h-d}})(y_2 + K)] = [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)],$$

with $d > 1$, then we would have

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h-1}})(y_2 + K) = [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_h})(y_2 + K)].$$

For $h = s$ this means we must show that

$$[A] \subset [P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K)]$$

for all distinct $i_1, \cdots, i_{s-1}$. This will be proved, however, if we show that

$$(54) \qquad (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_3 + K_1) \not\equiv 0 \quad [A],^{34}$$

for the first member here is in

---

[33] See (10).

[34] K_j is the j-th derivative of K.

$$[P]\cdot[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K)].^{35}$$

The relation (54) may be established in the same way in which (37) was disproved.

From (54) it easily follows that

$$(y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K) \not\equiv 0 \qquad (\Sigma_0).$$

For $h < s$ we proceed by induction. Assume that

$$(55) \qquad \begin{aligned}&(y - \eta_{i_1}) \cdots (y - \eta_{i_{j-1}})(y - \eta_{i_{j+1}}) \cdots (y - \eta_{i_{h+1}})(y_2 + K)\\ &\not\equiv 0 \ [A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{h+1}})(y_2 + K)]\end{aligned}$$

for all distinct $i, \cdots, i_{h+1}$ and all $j = 1, \cdots, h + 1$. (For $h = s - 1$ this has just been proved.)

Suppose, say, that

$$\begin{aligned}&(y - \eta_1) \cdots (y - \eta_{h-1})(y_2 + K)\\ &\equiv 0 \ [A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_h)(y_2 + K)].\end{aligned}$$

Then, for t sufficiently large,

$$\begin{aligned}&(y - \eta_1) \cdots (y - \eta_{h-1})(y - \eta_{h+1})^t(y_2 + K)\\ &\equiv 0 \ [A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_{h+1})(y_2 + K)].\end{aligned}$$

Combining this congruence with

$$\begin{aligned}&(y - \eta_1) \cdots (y - \eta_{h+1})(y_2 + K)\\ &\equiv 0 \ [A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_{h+1})(y_2 + K)],\end{aligned}$$

we see that

$$\begin{aligned}&(y - \eta_1) \cdots (y - \eta_{h-1})(y - \eta_{h+1})(y_2 + K)\\ &\equiv 0 \ [A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_{h+1})(y_2 + K)].\end{aligned}$$

This contradicts (55), and completes the first part of the proof.

As for the second part, let Σ be a differential ideal between Σ_0 and $\{\Sigma_0\}$, distinct from $\{\Sigma_0\}$. We shall complete the proof by showing that Σ coincides with some ideal

$$(56) \qquad [A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_g})(y_2 + K)].$$

---

[35] The first member of (54) is in $[P]$ because it is in $[y_1]$ and $y_1 \in [P]$; the first member of (54) is in $[A, (y^2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K)]$ because $((y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K))_1 = ((y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}}))'y_1(y_2 + K) + (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K_1) \equiv (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K_1)$, (Σ_0). Moreover, the ideals $[P]$ and $[A, (y_2 + K)^2, (y - \eta_{i_1}) \cdots (y - \eta_{i_{s-1}})(y_2 + K)]$ are separated, because P and $y_2 + K$ have no solution in common.

Let q be the smallest integer for which some ideal (56) with $g = q$ is included in Σ. q is an integer from 1 to s. Suppose, say, that

$$\Lambda \subseteq \Sigma,$$

where

$$\Lambda = [A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_q)(y_2 + K)],$$

but that

$$[A, (y_2 + K)^2, (y - \eta_1) \cdots (y - \eta_{i-1})(y - \eta_{i+1}) \cdots (y - \eta_q)(y_2 + K)] \not\subseteq \Sigma,$$

$i = 1, \cdots, q$. We shall prove that $\Sigma = \Lambda$.

Assume to the contrary that $F \in \Sigma$, $F \notin \Lambda$. Since $F \in \{\Sigma_0\} = [A, y_2 + K]$, we may write

$$F = C_0 A + C_1(y_2 + K) + \cdots + C_t(y_{t+1} + K_{t-1}). \tag{57}$$

Let D_t be the remainder of C_t with respect to $y_2 + K$. Since $(y_{i+1} + K_{i-1})(y_{t+1} + K_{t-1}) \equiv 0$ (Λ), $i = 1, 2, \cdots$, we have, by (57),

$$F \equiv C_1(y_2 + K) + \cdots + C_{t-1}(y_t + K_{t-2}) + D_t(y_{t+1} + K_{t-1}), \quad (\Lambda). \tag{58}$$

D_t is a polynomial in y, y_1. Let E_t be the sum of the terms of D_t free of y_1. Since, by (51), $(y_1(y_2 + K))_{t-1} \equiv 0$ (Λ), it follows that

$$y_1(y_{t+1} + K_{t-1}) \equiv -(t-1)y_2(y_t + K_{t-2}) - \cdots - y_t(y_2 + K), \quad (\Lambda). \tag{59}$$

From this and (58) we see that

$$F \equiv C_1'(y_2 + K) + \cdots + C_{t-1}'(y_t + K_{t-2}) + E_t(y_{t-1} + K_{t-1}), \quad (\Lambda), \tag{60}$$

where the C_i' are new forms.

Let H_t be the remainder obtained on dividing E_t by $(y - \eta_1) \cdots (y - \eta_q)$. Since $((y - \eta_1) \cdots (y - \eta_q)(y_2 + K))_{t-1} \equiv 0$ (Λ),

$$\begin{aligned}(y - \eta_1) \cdots (y - \eta_q)(y_{t+1} + K_{t-1}) &\equiv -(t-1)((y - \eta_1) \cdots (y - \eta_q))_1(y_t + K_{t-1}) \\ &\quad - \cdots - ((y - \eta_1) \cdots (y - \eta_q))_{t-1}(y_2 + K), \quad (\Lambda).\end{aligned}$$

Hence, it follows from (60) that

$$F \equiv C_1''(y_2 + K) + \cdots + C_{t-1}''(y_t + K_{t-2}) + H_t(y_{t+1} + K_{t-1}), \quad (\Lambda), \tag{61}$$

where the C_i'' are new forms.

Treating the first $t - 1$ terms in the second member of (61) as we did the second member of (57), we obtain, after $t - 1$ more steps, a relation

$$F \equiv H_1(y_2 + K) + \cdots + H_t(y_{t+1} + K_{t-1}), \quad (\Lambda), \tag{62}$$

where each H_i is a form of order zero and degree less than q.

At least one H_i must be nonzero, else $F \in \Lambda$. We may (and do) suppose that $H_t \neq 0$.

From the congruence $y_1(y_2 + K) \equiv 0$ (Λ), we obtain relations

$$y_1^j(y_{j+1} + K_{j-1}) \equiv 0 \quad (\Lambda), \qquad j = 1, 2, \cdots. \tag{63}$$

Therefore, if we multiply (62) by y_1^{t-1} we find

$$H_t y_1^{t-1}(y_{t+1} + K_{t-1}) \equiv 0 \quad (\Sigma).$$

Now, by (59) and (63),

$$\begin{aligned} y_1^{j-1}(y_{j+1} + K_{j-1}) &\equiv -(j-1)y_1^{j-2}y_2(y_j + K_{j-2}) \\ &\equiv (j-1)y_1^{j-2}K(y_j + K_{j-2}), \quad (\Lambda), \end{aligned}$$

so that

$$y_1^{t-1}(y_{t+1} + K_{t-1}) \equiv (t-1)!K^{t-1}(y_2 + K), \quad (\Lambda).$$

Hence,

$$H_t K^{t-1}(y_2 + K) \equiv 0 \quad (\Sigma).$$

If L is the sum of the terms of K free of y_1, we find, since $y_1(y_2 + K) \equiv 0$ (Σ)

$$H_t L^{t-1}(y_2 + K) \equiv 0 \quad (\Sigma).$$

But $(y - \eta_1) \cdots (y - \eta_q)(y_2 + K) \equiv 0$ (Σ). Hence, if J is the greatest common divisor of $H_t L^{t-1}$ and $(y - \eta_1) \cdots (y - \eta_q)$,

$$J(y_2 + K) \equiv 0 \quad (\Sigma).$$

But L is prime to $(y - \eta_1) \cdots (y - \eta_q)$, for otherwise the ideals $[y - \eta_1], \cdots, [y - \eta_q]$, $[A, y_2 + K]$ would not be separated. Hence J is a divisor of H_t, is of degree less than q, and is a *proper* divisor of $(y - \eta_1) \cdots (y - \eta_q)$. This contradicts the definition of q as the least g for which Σ includes an ideal (56), and completes the proof.

The results just established show that the ideals between $[A]$ and $\{A\}$ can be organized into ascending *chains*, all of the same length s. That is, if $\Lambda_1, \cdots, \Lambda_{h-1}$ are intermediate ideals such that

$$[A] \subset \Lambda_1 \subset \cdots \subset \Lambda_{h-1} \subset \{A\}$$

(chain of length h), then there are $s - h$ other intermediate ideals which together with $[A], \Lambda_1, \cdots, \Lambda_{h-1}, \{A\}$ form a chain of length s. Moreover, there exist no chains of length exceeding s. This illustrates a phenomenon which we shall discuss in Part III.

Part III. Chains

13. Generalities

We return now to an abstract differential domain of integrity with basis theorem, which contains the rational numbers.

The $n + 1$ differential ideals $\sigma_0, \sigma_1, \cdots, \sigma_{n-1}, \sigma^*$ are said to form a *chain* if

$$\sigma_0 \subset \sigma_1 \subset \cdots \subset \sigma_{n-1} \subset \sigma^*. \tag{64}$$

n is called the *length* of the chain. We shall call σ_0 and σ^* the *end ideals*, $\sigma_1, \cdots, \sigma_{n-1}$ the *intermediate ideals*, of the chain. The ideals σ_{i-1}, σ_i (also, σ_{n-1}, σ^*) will be said to be *contiguous*. A set of ideals of the chain will be called *non-contiguous* if no two ideals of the set are contiguous. A chain is called a *principal* chain if no pair of contiguous ideals have a differential ideal properly between them. A principal chain may be expressed schematically by means of dashes:

$$\sigma_0 - \sigma_1 - \cdots - \sigma_{n-1} - \sigma^*.$$

The notation $\sigma - \tau$ means that $\sigma \subset \tau$, but there is no ω with $\sigma \subset \omega \subset \tau$. A chain with end ideals σ_0, σ^* is said to be a *refinement* of the chain (64) if each intermediate ideal of (64) is an intermediate ideal of the chain in question.

Let $\sigma_0, \sigma_1, \cdots, \sigma_{n-1}, \sigma^*$ be a principal chain, and let there be a second principal chain $\sigma_0, \sigma_1, \cdots, \sigma_{i-1}, \tau_i, \sigma_{i+1}, \cdots, \sigma_{n-1}, \sigma^*$ with $\sigma_i \neq \tau_i$, that is, schematically, let

(65)
$$\sigma_0 - \sigma_1 - \cdots - \sigma_{i-1} \begin{array}{c} \nearrow \sigma_i \\ \searrow \tau_i \end{array} \sigma_{i+1} - \cdots - \sigma_{n-1} - \sigma^*.$$

We shall say that the first chain is *flexible* at the intermediate ideal σ_i, and that the second chain is obtained from the first by a *flex*.

The first theorem on chains that we shall state is that of Jordan-Hölder-Schreier. A differential domain of integrity $\mathfrak{D}$ forms a group under addition. As a set of operators on this group we may consider 1. multiplication by an element of $\mathfrak{D}$, 2. differentiation. The permissible invariant subgroups of this group with operators are precisely the differential ideals of $\mathfrak{D}$. Thus, we may state

THEOREM 7 (Jordan-Hölder-Schreier): *Two chains with the same end ideals have refinements of equal length, with differential rings of remainder classes which are isomorphic (in some order).*[36]

PROOF: See van der Waerden, *Moderne Algebra*, vol. 1, 2nd edition, §46, or H. Zassenhauss, *Lehrbuch der Gruppentheorie*, vol. 1, chapter 2, §5.

As a consequence of Theorem 7, we see that all principal chains with the same end ideals have the same length.

THEOREM 8 (Ore): *Two different principal chains with the same end ideals, can be obtained from each other by a finite number of successive flexes.*

PROOF: See O. Ore, *On the foundations of abstract algebra.* I, Annals of Mathematics, vol. 36 (1935), pp. 406–434.

14. The exponent of a principal chain

By the *exponent* of the chain (64) we shall mean $l_{\sigma^*\sigma_0}$. We shall be concerned with principal chains, exclusively. Of course, if $\sigma^* \not\subseteq \{\sigma_0\}$, then the chain (64) has exponent ∞, so we may limit ourselves to chains for which $\sigma^* \subseteq \{\sigma_0\}$.

[36] By the differential ring of remainder classes of a chain (64) is meant the differential rings $\sigma_1/\sigma_0, \sigma_2/\sigma_1, \cdots, \sigma^*/\sigma_{n-1}$.

Our chief result may be stated as follows.

THEOREM 9: Let

$$\sigma_0 - \sigma_1 - \cdots - \sigma_{n-1} - \sigma^*, \qquad \sigma^* \subseteq \{\sigma_0\}, \tag{66}$$

be a principal chain which is flexible at r non-contiguous intermediate ideals.[37] *Then the exponent of the chain* (66) *does not exceed* $n - r + 1$.

As a consequence of Theorem 9, we see that if a differential ideal σ has finite exponent p, then there is a chain, with end ideals σ and $\{\sigma\}$, of length $p - 1$.

PROOF: We begin by showing that, if $\sigma_0 - \sigma_1$, and $\sigma_1 \subseteq \{\sigma_0\}$, then $l_{\sigma_1\sigma_0} = 2$. Indeed, let $l_{\sigma_1\sigma_0} > 2$. Then there is a $b \in \sigma_1$ such that $b^2 \notin \sigma_0$. Hence

$$\sigma_0 \subset [\sigma_0, b^2] \subseteq \sigma_1.$$

Since $\sigma_0 - \sigma_1$, we see that $\sigma_1 = [\sigma_0, b^2]$. In particular, $b \in [\sigma_0, b^2]$, so that

$$b \equiv A \ (\sigma_0),$$

where A is a form in b with coefficients in $\mathfrak{D}$, which has no term of degree less than 2. Using the iterative process introduced in the paper cited under footnote 19, we see that $b \equiv 0 \ (\sigma_0)$. This contradiction shows that $l_{\sigma_1\sigma_0} = 2$.

We now show that the exponent of the principal chain (66) does not exceed $n + 1$.

For chains of length 1 we have just given the proof. Let $n > 1$, and assume the result for chains of length less than n.

By the assumption, $\sigma^{*^n} \subseteq \sigma_1$. Suppose $\sigma^{*^{n+1}} \not\subseteq \sigma_0$. Then

$$\sigma_0 \subset (\sigma_0, \sigma^{*^{n+1}}) \subseteq \sigma_1.$$

Since $\sigma_0 - \sigma_1$, we see that $\sigma_1 = (\sigma_0, \sigma^{*^{n+1}})$. Hence

$$\begin{aligned}\sigma^{*^{n+1}} = \sigma^*\sigma^{*^n} &\subseteq \sigma^*\sigma_1 = \sigma^*(\sigma_0, \sigma^{*^{n+1}}) \\ &\subseteq (\sigma_0, \sigma^{*^{n+2}}) \subseteq (\sigma_0, \sigma^{*^2}\sigma_1) \\ &\subseteq (\sigma_0, (\sigma_0, \sigma^{*^{n+3}})) = (\sigma_0, \sigma^{*^{n+3}}) \\ &\subseteq \cdots \subseteq (\sigma_0, \sigma^{*^{2n}}) \subseteq (\sigma_0, \sigma_1^2) \subseteq \sigma_0.\end{aligned}$$

This contradiction proves our statement.

We are now ready to complete the proof of our theorem. What we have just shown is that the theorem is valid for $r = 0$. We proceed by induction. Let $r > 0$ and suppose the theorem holds for fewer than r non-contiguous ideals.

Let σ_i be the intermediate ideal of (66) of lowest subscript, at which the chain is flexible. Then we have, with $\sigma_i \neq \tau_i$, the situation depicted by (65). Now

$$\sigma_i - \sigma_{i+1} - \cdots - \sigma_{n-1} - \sigma^*$$

[37] Clearly, r is an integer between 0 and the greatest integer in $n/2$, inclusive.

and

$$\tau_i - \sigma_{i+1} - \cdots - \sigma_{n-1} - \sigma^*$$

are both principal chains of length $n - i$, flexible at $r - 1$ intermediate ideals. Hence, by the induction assumption, $\sigma^{*^{n-i-r+2}} \subseteq \sigma_i$, $\sigma^{*^{n-i-r+2}} \subseteq \tau_i$. But $\sigma_{i-1} \subseteq \sigma_i \cap \tau_i \subset \sigma_i$ and $\sigma_{i-1} - \sigma_i$, so that $\sigma_{i-1} = \sigma_i \cap \tau_i$. Therefore, $\sigma^{*^{n-i-r+2}} \subseteq \sigma_{i-1}$.

Suppose, now, that $\sigma^{*^{n-r+1}} \nsubseteq \sigma_0$. Let σ_k be the σ_j of least subscript, such that $\sigma^{*^{n-j-r+1}} \subseteq \sigma_j$. Then $1 \leqq k \leqq i - 1$ and

$$\sigma^{*^{n-k-r+1}} \subseteq \sigma_k, \qquad \sigma^{*^{n-k-r+2}} \nsubseteq \sigma_{k-1}, \qquad \sigma^{*^{n-k-r+3}} \subseteq \sigma_{k-1}.$$

Hence

$$\sigma_{k-1} \subset (\sigma_{k-1}, \sigma^{*^{n-k-r+2}}) \subseteq \sigma_k,$$

so that, since $\sigma_{k-1} - \sigma_k$, $\sigma_k = (\sigma_{k-1}, \sigma^{*^{n-k-r+2}})$. Thus,

$$\begin{aligned} \sigma^{*^{n-k-r+2}} &= \sigma^*\sigma^{*^{n-k-r+1}} \subseteq \sigma^*\sigma_k \\ &= \sigma^*(\sigma_{k-1}, \sigma^{*^{n-k-r+2}}) \\ &\subseteq (\sigma_{k-1}, \sigma^{*^{n-k-r+3}}) = \sigma_{k-1}. \end{aligned}$$

This contradiction completes the proof of Theorem 9.

EXAMPLE 8. Consider the chain

$$[y^{n+1}, y_1], [y^n, y_1], \cdots, [y^2, y_1], [y]$$

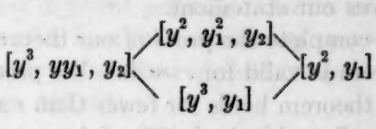
in any differential ring $\mathfrak{F}\{y\}$. It is easy to prove that this is a principal chain, flexible at none of its intermediate ideals. The length is n and the exponent is obviously $n + 1$. Thus, the bound on the exponent given by Theorem 9 is the exponent itself.

The necessity of the non-contiguity condition in Theorem 9 is shown by

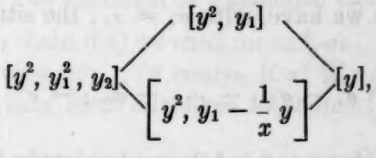
EXAMPLE 9. Let $\mathfrak{F}$ be a differential field of characteristic zero, which contains an element x with derivative 1. Consider the chain

$$[y^3, yy_1, y_2], [y^2, y_1^2, y_2], [y^2, y_1], [y]$$

in $\mathfrak{F}\{y\}$. This is a principal chain of length 3. Since



and



the chain is flexible at two intermediate ideals. Yet its exponent is 3.

Appendix. General Solutions and Resolvents

Liquidating an obligation incurred in Part I, we give here an abstract, purely algebraic treatment of the subject matter of Chapter 2 of Ritt's book, *Differential equations from the algebraic standpoint.*[38] We shall make free reference to this book, designating it by (*R*), and shall give proofs only when they differ from those in (*R*).

1*. The general solution of a form

We work in an abstract differential field $\mathfrak{F}$ of characteristic zero. The letters u_i, y_i and w will denote unknowns.

We establish the following fundamental lemma as a substitute for the existence theorem for differential equations.

Lemma 1*: Let A be an algebraically irreducible form in $\mathfrak{F}\{y_1, \cdots, y_n\}$ of positive class. Let $B_1, \cdots, B_t$ be nonzero forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$, reduced with respect to A. Then $B_1 \cdots B_t \notin \{A\}$.

Remark: Lemma 1*, together with the analog to the Hilbert-Netto theorem, shows that every algebraically irreducible form of positive class has a *regular* solution, that is, a solution for which the separant S and initial I of A do not vanish. For otherwise IS would be in $\{A\}$.

Proof: Let m be the class of A, r the order of A in y_m. Suppose $B_1 \cdots B_t \in \{A\}$, that is, $(B_1 \cdots B_t)^p \in [A]$, for some positive integer p. Now, A and $(B_1 \cdots B_t)^p$ are relatively prime. Hence there is a nonzero form C, either of class less than m, or of class m and of order less than r in y_m, such that

$$C = DA + E(B_1 \cdots B_t)^p,$$

where D and E are forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$. $C \in [A]$, so that we may write, for suitable q and Q_i,

$$C = Q_0A + Q_1A_1 + \cdots + Q_qA_q. \tag{1*}$$

Now, $A_i = Sy_{m,r+i} + T_i$, where T_i is a form of order less than $r + i$ in y_m. Hence, if in (1*) we consider the forms as polynomials in the letters y_{ij}, and let

$$y_{m,r+i} = -\frac{T_i}{S}, \qquad i = 1, 2, \cdots,$$

we see that S^hC is divisible by A for a sufficiently large integer h. Since neither S nor C is so divisible, we have a contradiction, completing the proof.

We now are in a position to follow (*R*).

Let $A \in \mathfrak{F}\{y_1, \cdots, y_n\}$ be algebraically irreducible and of class n. Let S and I be the separant and initial, respectively, of A. Let Σ_1 be the set of all forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$ which vanish for all regular solutions of A, or what is the same thing, let $\Sigma_1 = \{A\}:IS$.

[38] American Mathematical Society Colloquium Publications, vol. 14.

Following (R), we can show that Σ_1 *is a prime differential ideal.* The manifold of Σ_1 is called the *general solution* of A.

The following three results are proved as in (R):

$\Sigma_1 = \{A\}:S$;

Σ_1 *is an essential prime differential ideal in the decomposition of* $\{A\}$; *the other essential ideals contain* S;

Σ_1 *is independent of the order of the unknowns.*

2*. The resolvent of a prime differential ideal

Let Σ be a perfect differential ideal in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Let

$$A_1, \cdots, A_r \tag{2*}$$

be a basic set of Σ. We denote the separant and initial of A_i, $i = 1, \cdots, r$, by S_i and I_i, respectively. Let $\Sigma' = \{A_1, \cdots, A_r\}:I_1S_1 \cdots I_rS_r$, that is, let Σ' be the set of all forms in $\mathfrak{F}\{y_1, \cdots, y_n\}$ which vanish for all regular solutions of the basic set (2*).[39] Clearly, $\Sigma \subseteq \Sigma'$.

Following (R), we see that *if* Σ *is prime, then* $\Sigma = \Sigma'$. In particular (2*) has regular solutions. Also, Σ *is the only prime differential ideal with basic set* (2*).

Let $\mathfrak{C}$ be the set of all constants (that is, elements with vanishing derivative) in $\mathfrak{F}$. $\mathfrak{C}$ itself is a differential field of characteristic zero.

We shall need the following

LEMMA 2*: $\eta_1, \cdots, \eta_n \in \mathfrak{F}$ *are linearly dependent over* $\mathfrak{C}$ *if and only if*

$$\begin{vmatrix} \eta_1 & \cdots & \eta_n \\ \eta_{11} & \cdots & \eta_{n1} \\ \cdots & \cdots & \cdots \\ \eta_{1,n-1} & \cdots & \eta_{n,n-1} \end{vmatrix} = 0. \tag{3*}$$

PROOF: If $\sum_{\nu=1}^{n} c_\nu \eta_\nu = 0$, where the c_ν are in $\mathfrak{C}$ and not all 0, then $\sum_{\nu=1}^{n} c_\nu \eta_{\nu i} = 0$, $i = 0, 1, \cdots, n-1$, so that (3*) must hold. Conversely, let (3*) hold. If $n = 1$, the lemma is obviously valid. Let $n > 1$ and suppose the lemma holds for fewer than n elements η_ν. Let

$$\mu_\nu = (-1)^{n+\nu} \begin{vmatrix} \eta_1 & \cdots & \eta_{\nu-1} & \eta_{\nu+1} & \cdots & \eta_n \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \eta_{1,n-2} & \cdots & \eta_{\nu-1,n-2} & \eta_{\nu+1,n-2} & \cdots & \eta_{n,n-2} \end{vmatrix}, \qquad \nu = 1, \cdots, n.$$

If $\mu_n = 0$, the lemma for $n-1$ shows that the η_ν are linearly dependent over $\mathfrak{C}$. Assume that $\mu_n \neq 0$. By (3*),

$$\sum_{\nu=1}^{n} \eta_{\nu i} \mu_\nu = 0, \qquad i = 0, 1, \cdots, n-1.[40] \tag{4*}$$

[39] A solution of (2*) is called *regular* if it does not annul $I_1S_1 \cdots I_rS_r$.

[40] For $i = n-1$, the first member of (4*) is the development of the determinant in (3*) by the minors of the elements of the last row. For $i < n-1$, (4*) is a statement of the vanishing of certain determinants with two identical rows.

Differentiating the first $n-1$ equations (4*), we find

$$\sum_{\nu=1}^{n} \eta_{\nu,i+1}\mu_\nu + \sum_{\nu=1}^{n} \eta_{\nu i}\mu_{\nu 1} = 0, \qquad i = 0, 1, \cdots, n-2.$$

Taking into account the last $n-1$ equations (4*), we see that $\sum_{\nu=1}^{n} \eta_{\nu i}\mu_{\nu 1} = 0$, $i = 0, 1, \cdots, n-2$. These equations may be written

$$\sum_{\nu=1}^{n-1} \eta_{\nu i}\mu_{\nu 1} = -\eta_{ni}\mu_{n1}, \qquad i = 0, 1, \cdots, n-2.$$

Solving these equations in $\mu_{11}, \cdots, \mu_{n-1,1}$ by Cramer's rule, we obtain

$$\mu_{\nu 1} = \frac{\mu_{n1}}{\mu_n}\mu_\nu, \qquad \nu = 1, \cdots, n-1.$$

Hence $(\mu_\nu/\mu_n)_1 = (\mu_n\mu_{\nu 1} - \mu_{n1}\mu_\nu)/\mu_n^2 = 0$, so that $\mu_\nu/\mu_n = c_\nu$, $\nu = 1, \cdots, n$, where the c_ν are constants, not all zero ($c_n = 1$). Using (4*) with $i = 0$, we see that $\sum_{\nu=1}^{n} c_\nu\eta_\nu = 0$, which completes the proof.

LEMMA 3*: *Let $\mathfrak{F}$ contain a non-constant element ξ. Let A be a nonzero form in $\mathfrak{F}\{y_1, \cdots, y_n\}$. Then there are elements $\eta_1, \cdots, \eta_n \in \mathfrak{F}$ such that $A \neq 0$ when $y_1 = \eta_1, \cdots, y_n = \eta_n$.*

PROOF: It suffices to consider the case $n = 1$, for the cases $n > 1$ follow by induction. Let $A \in \mathfrak{F}\{y\}$ be of order p, and let $Y_j = \sum_{i=0}^{p} c_i(\xi^i)_j$, $j = 0, 1, \cdots, p$, where the c_i are new unknowns.[41] The Jacobian of the $p+1$ functions Y_j of the $p+1$ variables c_i is

$$\omega = \begin{vmatrix} 1 & \xi & \xi^2 & \cdots & \xi^p \\ 0 & \xi_1 & (\xi^2)_1 & \cdots & (\xi^p)_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \xi_p & (\xi^2)_p & \cdots & (\xi^p)_p \end{vmatrix}.$$

If ω were 0, ξ would be algebraic over $\mathfrak{C}$ (by Lemma 2*) and would therefore be a constant, so that $\omega \neq 0$. Hence $Y_0, \cdots, Y_p$, considered as polynomials in $c_0, \cdots, c_p$, are algebraically independent.[42] Therefore, if we substitute $Y_0, Y_1, \cdots, Y_p$ in A for $y, y_1, \cdots, y_p$, respectively, A will go over into a nonzero polynomial C in the c_i. Choose integers $\gamma_0, \gamma_1, \cdots, \gamma_p$ such that $C \neq 0$ when $c_i = \gamma_i$, $i = 0, 1, \cdots, p$. Then the element $\eta = \sum_{i=0}^{p} \gamma_i\xi^i \in \mathfrak{F}$ leaves A nonzero when substituted for y.

Using Lemma 3*, we can follow the proof of §25 in (*R*) to obtain the following result.

Let Σ be a non-trivial prime differential ideal in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$ with arbitrary unknowns $u_1, \cdots, u_q$ (q may be 0.) If $\mathfrak{F}$ contains a non-constant, then there are elements $\mu_1, \cdots, \mu_p \in \mathfrak{F}$, and a nonzero form $G \in \mathfrak{F}\{u_1, \cdots, u_q\}$, such that if

---

[41] $(\xi^i)_j$ is the j-th derivative of ξ^i.

[42] See O. Perron, *Algebra*, vol. 1, Berlin, 1932, p. 134.

(5*) $\bar{u}_1, \cdots, \bar{u}_q, y_1', \cdots, y_p'$ *and* $\bar{u}_1, \cdots, \bar{u}_q, y_1'', \cdots, y_p''$

are two distinct solutions of Σ, *then either* $G(\bar{u}_1, \cdots, \bar{u}_q) = 0$, *or*

$$\mu_1(y_1' - y_1'') + \cdots + \mu_p(y_p' - y_p'') \neq 0.$$

Taking over the proof in §26 of (*R*), we obtain:

Let $\mathfrak{F}$ *be any differential field of characteristic zero. Let* Σ *be a nontrivial prime differential ideal in* $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$ *with arbitrary unknowns* $u_1, \cdots, u_q$. *We suppose that* $q > 0$. *Then there are forms*

$$G, M_1, \cdots, M_p \in \mathfrak{F}\{u_1, \cdots, u_p\},$$

with $G \neq 0$, *such that if* (5*) *are two distinct solutions of* Σ, *then either* $G(\bar{u}_1, \cdots, \bar{u}_q) = 0$ *or*

$$M_1(\bar{u}_1, \cdots, \bar{u}_q)\cdot(y_1' - y_1'') + \cdots + M_p(\bar{u}_1, \cdots, \bar{u}_q)\cdot(y_p' - y_p'') \neq 0.$$

Henceforth we shall deal with a non-trivial prime differential ideal Σ in $\mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with arbitrary unknowns $u_1, \cdots, u_q$, and we shall assume that either $\mathfrak{F}$ contains non-constant elements, or the arbitrary unknowns u_i really exist. Then there exists a triad of forms G, P, $Q \in \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}$, with G free of the y_i, and not 0, and with $P \notin \Sigma$, such that if (5*) are two distinct solutions of Σ for which $GP \neq 0$, then

$$\frac{Q(\bar{u}_1, \cdots, \bar{u}_q, y_1', \cdots, y_p')}{P(\bar{u}_1, \cdots, \bar{u}_q, y_1', \cdots, y_p')} \neq \frac{Q(\bar{u}_1, \cdots, \bar{u}_q, y_1'', \cdots, y_p'')}{P(\bar{u}_1, \cdots, \bar{u}_q, y_1'', \cdots, y_p'')}.$$

Following (*R*), we form the ideal

$$\Omega = \{\Sigma, Pw - Q\}:P \subset \mathfrak{F}\{u_1, \cdots, u_q, w, y_1, \cdots, y_p\}.$$

Ω is prime, with arbitrary unknowns $u_1, \cdots, u_q$. Further,

$$\Sigma = \Omega \cap \mathfrak{F}\{u_1, \cdots, u_q, y_1, \cdots, y_p\}.$$

Let

(6*) $$A, A_1, \cdots, A_p$$

be a basic set for Ω, introducing $w, y_1, \cdots, y_p$ in succession. We take, as we may, A algebraically irreducible.

We shall show that *each* A_i *is of order zero in* y_i *and is linear in* y_i. The equation $A = 0$ is called a *resolvent* of Σ.

Suppose that not every A_i is as described. Let A_j be the first A_i which either is of positive order or is not linear in y_i, and suppose that the order of A_j in y_j is r, with r positive.

Let U be the remainder of $ISI_1S_1 \cdots I_pS_pGP$ with respect to (6*), where the I's and J's are the initials and separants of the forms of the basic set (6*). Then U has the properties: 1. The order of U in w is not greater than that of A; 2. The order of U in y_i, $i = 1, \cdots, j$, is not greater than that of A_i; 3. U is free of $y_{j+1}, \cdots, y_p$; 4. $U \notin \Omega$.

Of all the forms in (U, Ω) which have the properties 1–4 listed for U, let V be one of least rank in y_j.

We shall prove that the order of V in y_j is less than r. Suppose it equals r. Then the degree of V in y_{jr} is no higher than that of A_j. Let J be the initial of V. Then $J \notin \Omega$.

For an appropriate integer m, we have

$$J^m A_j = EV + F,$$

where E is a nonzero form of lower degree than A_j in y_{jr}, and where F has the properties 1–3 and is of lower degree than V in y_{jr}. Hence $F \in \Omega$, so that $EV \in \Omega$, whence $E \in \Omega$.

Let $E = H_0 + H_1 y_{jr} + \cdots + H_t y_{jr}^t$, where $H_t \not\equiv 0$, and each H_i is free of $y_{j+1}, \cdots, y_p$ and is of order less than r in y_j. The initial of $J^m A_j$ is identical with that of EV. Hence $H_t \notin \Omega$.

It is easy to see that there exist positive integers $a, a_1, \cdots, a_{j-1}$ such that

$$I^a I_1^{a_1} \cdots I_{j-1}^{a_{j-1}} H_i \equiv H_i' \quad (\Omega), \qquad i = 0, 1, \cdots, t,$$

where each H_i' is reduced with respect to the ascending set $A, A_1, \cdots, A_{j-1}$. Since $H_t \notin \Omega$, we have $H_t' \not\equiv 0$. Thus, $H_0' + H_1' y_{jr} + \cdots + H_t' y_{jr}^t$ is a nonzero form in Ω, reduced with respect to the basic set (6*). This contradiction shows that the order of V in y_j is less than r.

Now let

$$(7^*) \qquad \bar{u}_1, \cdots, \bar{u}_q, \bar{w}, \bar{y}_1, \cdots, \bar{y}_p$$

be a solution of Ω for which V (and consequently $ISI_1S_1 \cdots I_pS_pGP$) does not vanish. For $\bar{u}_1, \cdots, \bar{u}_q, \bar{w}, \bar{y}_1, \cdots, \bar{y}_{j-1}$, the forms $A_1, \cdots, A_{j-1}$ all vanish, A_j becomes a nonzero form $\bar{A}_j$ in y_j of order r, A_i, $i = j+1, \cdots, p$, becomes a form $\bar{A}_i$ of order zero and linear in y_i, and V becomes a form $\bar{V}$ in y_j of order less than r. It follows, by Lemma 1*, that $(y_j - \bar{y}_j)\bar{V}$ does not hold $\bar{A}_j$.

Let $\tilde{y}_j$ be a solution of $\bar{A}_j$ for which $(y_j - \bar{y}_j)\bar{V}$ does not vanish. When $y_j = \tilde{y}_j$, $\bar{A}_i$, $i = j+1, \cdots, p$, becomes a linear form of order zero in y_i alone. The new forms $\bar{A}_i$, $i = j+1, \cdots, p$, may be solved to obtain a solution $\tilde{y}_{j+1}, \cdots, \tilde{y}_p$.

Then $\bar{u}_1, \cdots, \bar{u}_q, \bar{w}, \bar{y}_1, \cdots, \bar{y}_{j-1}, \tilde{y}_j, \cdots, \tilde{y}_p$ is a solution of Ω for which GP does not vanish, is distinct from (7*), but has the same $w = \bar{w}$ and the same $u_i = \bar{u}_i$, $i = 1, \cdots, q$. This contradicts the fundamental property of the triad G, P, Q, and shows that A_j is of order zero in y_j.

That A_j is of degree unity in y_j may be proved as in (R). This completes the proof.

The rest of Chapter 2 in (R), except for §33, may now be taken over, word for word.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

HANKEL AND OTHER EXTENSIONS OF DIRICHLET'S SERIES

BY R. E. GREENWOOD

(Received December 14, 1939)

1. Introduction

Series of the type

$$f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s), \tag{1.1}$$

where (a) s is the complex variable, $s = \sigma + it$, σ and t real;
(b) the numbers a_n are all complex constants;
(c) $\lambda_1 < \lambda_2 < \lambda_3 < \cdots < \lambda_n < \cdots$

$$\lim_{n\to\infty} \lambda_n = \infty$$

are known as Dirichlet series, and their theory has been considered by numerous investigators. Widder[1] has generalized these series by the consideration of a Stieltjes integral.

If we consider series of the type

$$g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s) \tag{1.2}$$

it may be possible to establish convergence properties, subject to certain restrictions on the function $G(z)$, for the series (1.2) similar to the well known convergence properties of the Dirichlet series (1.1).

In particular, this method of generalizing Dirichlet series was suggested from a consideration of series of the type

$$h(s) = \sum_{n=1}^{\infty} a_n (i\lambda_n s)^{\frac{1}{2}} H_\nu^{(1)}(i\lambda_n s), \tag{1.3}$$

where $H_\nu^{(1)}(z)$ is the Hankel function of the first kind of order ν and is related to the more familiar Bessel functions $J_\nu(z)$ and $J_{-\nu}(z)$ by the relation

$$H_\nu^{(1)}(z) = \frac{J_{-\nu}(z) - \exp(-\nu\pi i) J_\nu(z)}{i \sin \nu\pi}. \tag{1.4}$$

The function $H_\nu^{(1)}(z)$ has the integral representation[2] for $\Re(\nu) \geqq -\frac{1}{2}$

[1] D. V. Widder, "A Generalization of Dirichlet Series," Transactions of the American Mathematical Society, Vol. 31, p. 694f. (1929).

[2] G. N. Watson, "The Theory of Bessel Functions," page 168.

$$(1.5)\quad H_\nu^{(1)}(z) = \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \frac{\exp\left[i\left(z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right]}{\Gamma(\nu + \frac{1}{2})} \int_0^\infty \exp(-u)u^{\nu-\frac{1}{2}}\left(1 + \frac{iu}{2z}\right)^{\nu-\frac{1}{2}} du.$$

If we replace z by iz and define

$$(1.6)\quad M_\nu(z) = \int_0^\infty \exp(-u)u^{\nu-\frac{1}{2}}\left(1 + \frac{u}{2z}\right)^{\nu-\frac{1}{2}} du$$

$$(1.7)\quad A_\nu = \frac{\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp\left[-i\left(\frac{\nu\pi}{2} + \frac{\pi}{4}\right)\right]}{\Gamma(\nu + \frac{1}{2})}$$

we see that

$$(1.8)\quad (iz)^{\frac{1}{2}} H_\nu^{(1)}(iz) = A_\nu \exp(-z) M_\nu(z).$$

Therefore the series (1.3) can also be written as

$$(1.9)\quad h(s) = \sum_{n=1}^\infty a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s)$$

and thus is formally identifiable as a series of type (1.2).

For convenience we shall call series of type (1.2) modified exponential series, and the particular series of type (1.9) Hankel series. Moreover we shall suppose that we are given a fixed λ_n sequence satisfying

(a) $$0 < \lambda_1 < \lambda_2 < \lambda_3 < \cdots < \lambda_n < \cdots,$$

(b) $$\lim_{n\to\infty} \lambda_n = \infty.$$

The symbol $K(a)$ will be used to indicate a positive constant, depending perhaps on a.

2. Lemmas Concerning the Function $G(z)$

In order to establish convergence properties for the modified exponential series (1.2) we shall assume that the function $G(z)$ satisfies the following properties (which properties are understood to be assumed whenever $G(z)$ is mentioned hereafter):

PROPERTY 1. *$G(z)$ is analytic for $\Re(z) > 0$.*

PROPERTY 2. *$G(z)$ is bounded for $\Re(z) \geqq z_0 > 0$, i.e. say*

$$|G(z)| < K(z_0) \quad \text{for} \quad \Re(z) \geqq z_0 > 0.$$

PROPERTY 3. *$G(z)$ has an asymptotic expansion for large z of the form*

$$(2.1)\quad G(z) \cong b_0 + \frac{b_1}{z} + \frac{b_2}{z^2} + \cdots,$$

where $b_0 \neq 0$.

On the basis of these properties we can establish a number of lemmas, which we shall state for the function $G(xs)$ where x is a real variable.

LEMMA 1. *There exists a constant τ such that for*

(a) $$x \geqq \lambda_1 > 0,$$

(b) $$\Re(s) \geqq \tau > 0,$$

we have

$$|G(xs)| < K_1(\tau)^3.$$

PROOF: This is merely a restatement of property 2.

LEMMA 2. *Given $\tau > 0$, there exists an integer $N_0(\tau)$ such that for*

(a) $$x \geqq \lambda_n, \qquad n \geqq N_0(\tau),$$

(b) $$\Re(s) \geqq \tau > 0,$$

we have

$$0 < K_2(\tau) \leqq |G(xs)|.$$

PROOF: Using the asymptotic expansion

$$G(xs) \cong b_0 + \frac{b_1}{xs} + \frac{b_2}{x^2 s^2} + \cdots, \tag{2.1}$$

and noting that $b_0 \neq 0$, we can get

$$|G(xs) - b_0| < \frac{B}{x\tau} \qquad \text{for large } x$$

where B is a constant, and hence we can clearly select x values and a constant $K_2(\tau)$ so that we obtain

$$0 < K_2(\tau) < |G(xs)|.$$

LEMMA 3. *Given $\lambda_1 > 0$, there exists a real number $\xi(\lambda_1)$ such that for*

(a) $$x \geqq \lambda_1 > 0,$$

(b) $$\Re(s) \geqq \xi > 0,$$

we have

$$0 < K_3(\tau) < |G(xs)|.$$

Proof is similar to that of lemma 2.

LEMMA 4. *For*

(a) $$\Re(s) \geqq \Re(s_0) = \tau > 0,$$

(b) $$x \geqq \lambda_n, \qquad n \geqq N_0(\tau),$$

---

[3] In reality K_1 depends also on λ_1, but we are considering the λ_n sequence as fixed and do not indicate this dependence.

we have

$$\left|\frac{G(xs)}{G(xs_0)}\right| \leqq K_4(\tau);$$

as follows at once from combining lemmas 1 and 2.

Now the asymptotic expansion for $dG(z)/dz$ can be obtained from the asymptotic expansion,

$$G(xs) \cong b_0 + \frac{b_1}{xs} + \frac{b_2}{x^2 s^2} + \cdots, \tag{2.1}$$

by termwise differentiation,[4] thereby leading to

$$\frac{\partial G(xs)}{\partial s} \cong -\frac{b_1}{xs^2} - \frac{2b_2}{x^2 s^3} - \cdots, \tag{2.2}$$

$$\frac{\partial G(xs)}{\partial x} \cong -\frac{b_1}{x^2 s} - \frac{2b_2}{x^3 s^2} - \cdots, \tag{2.3}$$

$$\frac{\partial^2 G(xs)}{\partial x \partial s} \cong \frac{b_1}{x^2 s^2} + \frac{4b_2}{x^3 s^3} + \cdots, \tag{2.4}$$

and from these expansions we obtain

LEMMA 5. *For $\Re(s) \geqq \tau > 0$ there exists an integer $N_1(\tau)$ such that for $x \geqq \lambda_n$, $n \geqq N_1(\tau)$ we have*

$$\left|\frac{\partial G(xs)}{\partial s}\right| < K_5(\tau),$$

$$\left|\frac{\partial G(xs)}{\partial x}\right| < K_6(\tau),$$

$$\left|\frac{\partial^2 G(xs)}{\partial x \partial s}\right| < K_7(\tau).$$

LEMMA 6. *For $\Re(s) \geqq \tau > 0$ there exists an integer $N_2(\tau)$ such that for $q \geqq p \geqq N_2(\tau)$ we have*

$$\int_{\lambda_p}^{\lambda_q} \left|\frac{\partial G(xs)}{\partial x}\right| dx \leqq K_8(\tau)\left[\frac{1}{\tau\lambda_p} - \frac{1}{\tau\lambda_q}\right] \leqq \frac{K_8(\tau)}{\tau\lambda_p}.$$

PROOF. This follows immediately from the asymptotic expansion (2.3). The inequality

$$\left|\frac{|f(z+h)| - |f(z)|}{h}\right| \leqq \frac{|f(z+h) - f(z)|}{|h|},$$

[4] Bochner, "Fourier Analysis," (planographed notes, Princeton University), pages 7 and 29; also Knopp, "Theorie und Anwendung der Unendlichen Reihen," third edition, p. 560.

shows that

$$\left|\frac{d|f(z)|}{dz}\right| \leqq \left|\frac{df(z)}{dz}\right|$$

and thus leads to

LEMMA 7. *For $\Re(s) \geqq \tau > 0$ and for $q \geqq p \geqq N_2(\tau)$ we have*

$$\left|\int_{\lambda_p}^{\lambda_q} \frac{\partial|G(xs)|}{\partial x}\,dx\right| \leqq \frac{K_8(\tau)}{\tau\lambda_p}.$$

LEMMA 8. *For $\Re(s) \geqq \Re(s_0) = \tau > 0$ and for $x \geqq \lambda_n$, $n \geqq \max[N_1(\tau), N_2(\tau)]$ we have*

$$\left|\frac{\dfrac{\partial G(xs_0)}{\partial x}}{G(xs_0)} - \frac{\dfrac{\partial G(xs)}{\partial x}}{G(xs)}\right| \leqq K_9(\tau)|s - s_0|.$$

PROOF. Writing the function as the integral of its own derivative, we have

$$\left|\frac{\dfrac{\partial G(xs_0)}{\partial x}}{G(xs_0)} - \frac{\dfrac{\partial G(xs)}{\partial x}}{G(xs)}\right| = \left|\int_{s_0}^{s} \frac{\dfrac{\partial^2 G(xz)}{\partial z\partial x}\,G(xz) - \dfrac{\partial G(xz)}{\partial x}\,\dfrac{\partial G(xz)}{\partial z}}{[G(xz)]^2}\,dz\right|$$

$$\leqq \left|\int_{s_0}^{s} \frac{K_7(\tau)K_1(\tau) + K_5(\tau)K_6(\tau)}{[K_2(\tau)]^2}\,dz\right| \leqq K_9(\tau)|s - s_0|$$

by applying lemmas 1, 2, and 5.

In particular, the function $M_\nu(z)$ defined by

$$M_\nu(z) = \int_0^\infty \exp(-u)u^{\nu-\frac{1}{2}}\left(1 + \frac{u}{2z}\right)^{\nu-\frac{1}{2}}du, \tag{1.6}$$

satisfies Properties 1, 2, and 3. For $M_\nu(z)$ is analytic both in z and in ν, and has the asymptotic expansion[5]

$$M_\nu(z) \cong \sum_{m=0}^{\infty} \frac{\Gamma(\nu + \frac{1}{2})\Gamma(\nu + \frac{1}{2} + m)}{m!\Gamma(\nu + \frac{1}{2} - m)(2z)^m}$$

$$\cong \Gamma(\nu + \tfrac{1}{2})\left[1 + \sum_{m=1}^{\infty} \frac{\{4\nu^2 - 1^2\}\{4\nu^2 - 3^2\}\cdots\{4\nu^2 - (2m-1)^2\}}{2^{3m}m!z^m}\right]$$

and thus, since $b_0 \neq 0$, properties 1 and 3 are satisfied. Property 2 can be established by taking an upper bound to the complex valued integral representation (1.6) for $\Re(z) \geqq z_0 > 0$.

Thus any convergence proof for the modified exponential series

$$g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)G(\lambda_n s) \tag{1.2}$$

[5] Watson, l.c., p. 198, and Bochner, l.c., page 15. Care must be taken in interpreting the notations in these references for the case $m = 0$.

which uses only properties 1, 2, and 3 of the function $G(z)$ shows at the same time that the corresponding Hankel series

$$h(s) = \sum_{n=1}^{\infty} a_n A_r \exp(-\lambda_n s) M_r(\lambda_n s) \tag{1.9}$$

satisfies the same type of convergence.

For convenience we state a familiar lemma.

LEMMA 9. Abel's lemma on partial summation. *We have identically*

$$\sum_{n=p}^{n=q} U_n V_n \equiv \sum_{n=p}^{n=q-1} \left[(V_n - V_{n+1}) \sum_{m=p}^{m=n} U_m \right] + V_q \sum_{n=p}^{n=q} U_n .$$

3. Convergence of the Modified Exponential Series

THEOREM 1. *If the modified exponential series* (1.2) *is convergent for* $s_0 = \sigma_0 + it_0$ *where* $\sigma_0 > 0$, *then it is convergent for any* s *satisfying* $\Re(s) \geqq \Re(s_0) = \sigma_0$.

PROOF: The proof of this theorem is contained in the more general theorem which follows.

THEOREM 2. *If the modified exponential series* (1.2) *is convergent for* $s_0 = \sigma_0 + it_0$ *where* $\sigma_0 > 0$, *then it is uniformly convergent throughout the angular sector about the point* s_0 *in the* s *plane defined by the inequality*

$$|\operatorname{am}(s - s_0)| \leqq \psi < \frac{\pi}{2}.$$

PROOF: We need only show that for s in the angular sector, given $\delta > 0$, we can find an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| < \delta.$$

In Abel's lemma on partial summation we identify

$$U_n = a_n \exp(-\lambda_n s_0) G(\lambda_n s_0),$$

$$V_n = \exp[-(s - s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)},$$

so that we have

$$\begin{aligned} \left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| &= \left| \sum_{n=p}^{n=q} U_n V_n \right| \\ &\leqq \sum_{n=p}^{n=q-1} \left| \exp[-(s - s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)} - \exp[-(s - s_0)\lambda_{n+1}] \frac{G(\lambda_{n+1} s)}{G(\lambda_{n+1} s_0)} \right| \\ &\qquad \cdot \left| \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right| \\ &\quad + \left| \exp[-(s - s_0)\lambda_q] \frac{G(\lambda_q s)}{G(\lambda_q s_0)} \right| \cdot \left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right|. \end{aligned} \tag{3.1}$$

By hypothesis, given an $\epsilon > 0$, there exists an integer n_1 such that for $q \geqq p \geqq n_1$ we have

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0)\right| < \epsilon.$$

We take $n_0 = \max[N_0(\tau), N_1(\tau), n_1]$, where N_0 is given by lemma 4 and N_1 by lemma 5. Then (3.1) becomes

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s)\right|$$

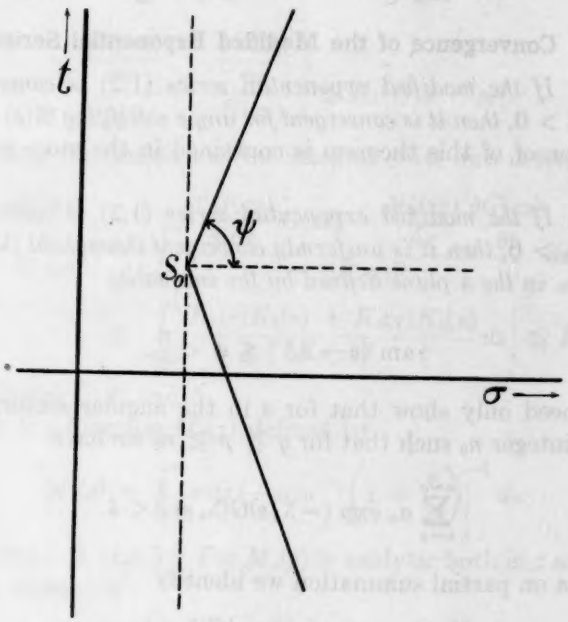


FIG. 1

$$(3.2) \quad \leqq \epsilon \sum_{n=p}^{n=q-1} \left|\exp[-(s-s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)} - \exp[-(s-s_0)\lambda_{n+1}] \frac{G(\lambda_{n+1} s)}{G(\lambda_{n+1} s_0)}\right| + \epsilon \left|\exp[-(s-s_0)\lambda_q] \frac{G(\lambda_q s)}{G(\lambda_q s_0)}\right|.$$

Writing the function as the integral of its own derivative, we have for $n \geqq n_0$

$$\left|\exp[-(s-s_0)\lambda_n] \frac{G(\lambda_n s)}{G(\lambda_n s_0)} - \exp[-(s-s_0)\lambda_{n+1}] \frac{G(\lambda_{n+1} s)}{G(\lambda_{n+1} s_0)}\right| \leqq \left|(s-s_0) \int_{\lambda_n}^{\lambda_{n+1}} \exp[-(s-s_0)x] \frac{G(xs)}{G(xs_0)}\, dx\right|$$

$$+\left|\int_{\lambda_n}^{\lambda_{n+1}} \exp\left[-(s-s_0)x\right]\frac{G(xs)}{G(xs_0)}\left\{\frac{\frac{\partial G(xs_0)}{\partial x}}{G(xs_0)}-\frac{\frac{\partial G(xs)}{\partial x}}{G(xs)}\right\}\right| \tag{3.3}$$

$$\leqq K_4(\sigma_0)\cdot|s-s_0|\cdot\left|\int_{\lambda_n}^{\lambda_{n+1}}\exp\left[-(s-s_0)x\right]dx\right|$$

$$+K_4(\sigma_0)K_9(\sigma_0)\cdot|s-s_0|\cdot\left|\int_{\lambda_n}^{\lambda_{n+1}}\exp\left[-(s-s_0)x\right]dx\right|$$

by application of lemmas 4 and 8, and hence

$$\leqq K_4(\sigma_0)(1+K_9(\sigma_0))\frac{|s-s_0|}{\Re(s-s_0)}\left[\exp\left[-\Re(s-s_0)\lambda_n\right]-\exp\left[-\Re(s-s_0)\lambda_{n+1}\right]\right].$$

Substituting this inequality in (3.2) we get

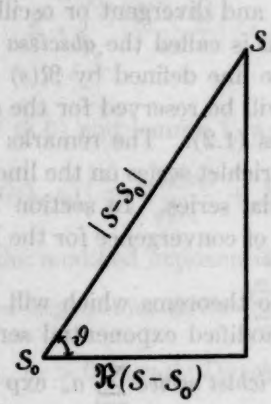


FIG. 2

$$\left|\sum_{n=p}^{n=q} a_n \exp\left(-\lambda_n s\right)G(\lambda_n s)\right|$$

$$\leqq \frac{\epsilon|s-s_0|}{\Re(s-s_0)}K_4(\sigma_0)[1+K_9(\sigma_0)]\left[\exp\left[-\Re(s-s_0)\lambda_p\right]-\exp\left[-\Re(s-s_0)\lambda_q\right]\right]$$

$$+\epsilon K_4(\sigma_0)\exp\left[-\Re(s-s_0)\lambda_q\right] \tag{3.4}$$

$$\leqq \frac{\epsilon|s-s_0|}{\Re(s-s_0)}K_4(\sigma_0)[1+K_9(\sigma_0)]\exp\left[-\Re(s-s_0)\lambda_p\right]$$

$$\leqq \epsilon K_{10}(\sigma_0)\frac{|s-s_0|}{\Re(s-s_0)}.$$

Now for am $(s-s_0)=\theta$, $|\theta|\leqq\psi<\frac{\pi}{2}$, we get

$$\frac{|s-s_0|}{\Re(s-s_0)}=\sec\theta\leqq\sec\psi,$$

and thus (3.4) becomes

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| \leqq \epsilon K_{10}(\sigma_0) \sec \psi,$$

which holds for all s in the angular sector and for $q \geqq p \geqq n_0$, and thus the required condition is satisfied if we specify that we choose $\epsilon = \delta/(K_{10}(\sigma_0) \sec \psi)$ and the proof is complete.

DEFINITION: We suppose that the modified exponential series is not identically zero. Then, as is the case with Dirichlet series, the series (1.2) may:

(a) Converge for all s, $\Re(s) > 0$

(b) Converge for no s, $\Re(s) > 0$

(c) Converge for some values of s, $\Re(s) > 0$.

In the latter case there is a real number α such that the series is convergent for all s satisfying $\Re(s) > \alpha$, and divergent or oscillatory for all s satisfying $0 < \Re(s) < \alpha$. The number α is called the *abscissa of convergence* of the modified exponential series, and the line defined by $\Re(s) = \alpha$ is called the *line of convergence*. The symbol α will be reserved for the abscissa of convergence of the modified exponential series (1.2). The remarks of Hardy and Riesz[6] relative to the convergence of a Dirichlet series on the line of convergence are applicable also to modified exponential series. In section 7 of this paper we consider a special case where the line of convergence for the Hankel series (1.9) contains at least one singular point.

We shall now prove two theorems which will show equi-convergence of the Dirichlet series and the modified exponential series.

THEOREM 3. *If the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ *converges for the point* $s_0 = \sigma_0 + it_0$, $\sigma_0 > 0$, *then the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0)$ *is also convergent.*

PROOF: In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s_0),$$
$$V_n = G(\lambda_n s_0),$$

and thus we get

$$\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0)$$

$$\equiv \sum_{n=p}^{n=q-1} [G(\lambda_n s_0) - G(\lambda_{n+1} s_0)] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) \right] + G(\lambda_q s_0) \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s_0). \tag{3.5}$$

[6] Hardy and Riesz, "The General Theory of Dirichlet's Series," page 9.

Now by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) \right| < \epsilon.$$

Now choose $n_1 = \max [n_0, N_2(\sigma_0)]$ where N_2 is given by lemma 6. We get for $q \geqq p \geqq n_1$

$$\left| \sum_{n=p}^{n=q-1} [G(\lambda_n s_0) - G(\lambda_{n+1} s_0)] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) \right] \right|$$

$$\leqq \sum_{n=p}^{n=q-1} \left| \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) \right| \int_{\lambda_n}^{\lambda_{n+1}} \left| \frac{\partial G(x s_0)}{\partial x} \right| dx \tag{3.6}$$

$$\leqq \epsilon \int_{\lambda_p}^{\lambda_q} \left| \frac{\partial G(x s_0)}{\partial x} \right| dx \leqq \frac{\epsilon K_3(\sigma_0)}{\sigma_0 \lambda_p} \leqq \frac{\epsilon K_3(\sigma_0)}{\sigma_0 \lambda_1}$$

on application of lemma 6.

On reducing (3.5) by use of (3.6) and lemma 1, we get, for $q \geqq p \geqq n_1$

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0) \right| \leqq \frac{\epsilon K_3(\sigma_0)}{\sigma_0 \lambda_1} + K_1(\sigma_0)\epsilon \leqq \epsilon K(\sigma_0),$$

and thus the convergence of the modified exponential series is established.

THEOREM 4. *If the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *converges for the point* $s_0 = \sigma_0 + it_0$, $\sigma_0 > 0$, *then the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_0)$ *is also convergent.*

PROOF. In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s_0) G(\lambda_n s_0),$$

$$V_n = \frac{1}{G(\lambda_n s_0)}.$$

Now, by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) G(\lambda_n s_0) \right| < \epsilon.$$

Now choose $n_1 = \max [N_0(\sigma_0), N_2(\sigma_0), n_0]$ where N_0 and N_2 are given by lemmas 2 and 6 respectively. Similar to the proof of the previous theorem, we get, for $q \geqq p \geqq n_1$

$$\left| \sum_{n=p}^{n=q-1} \left[\frac{1}{G(\lambda_n s_0)} - \frac{1}{G(\lambda_{n+1} s_0)} \right] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right] \right|$$

$$(3.7)\qquad \leqq \sum_{n=p}^{n=q-1} \left| \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right| \int_{\lambda_n}^{\lambda_{n+1}} \left| \frac{\partial\left(-\frac{1}{G(xs_0)}\right)}{\partial x} \right| dx$$

$$\leqq \epsilon \int_{\lambda_p}^{\lambda_q} \left| \frac{\frac{\partial G(xs_0)}{\partial x}}{[G(xs_0)]^2} \right| dx \leqq \frac{\epsilon K_3(\sigma_0)}{\sigma_0 \lambda_p [K_2(\sigma_0)]^2}$$

by application of lemma 2 and lemma 6.

Abel's lemma then takes the form

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s_0) \right|$$

$$\leqq \left| \sum_{n=p}^{n=q-1} \left[\frac{1}{G(\lambda_n s_0)} - \frac{1}{G(\lambda_{n+1} s_0)} \right] \left[\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right] \right|$$

$$+ \left| \frac{1}{G(\lambda_q s_0)} \right| \cdot \left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s_0) G(\lambda_m s_0) \right|$$

$$\leqq \frac{\epsilon K_3(\sigma_0)}{\sigma_0 \lambda_p [K_2(\sigma_0)]^2} + \frac{\epsilon}{K_2(\sigma_0)} \leqq \epsilon K(\sigma_0)$$

for $q \geqq p \geqq n_1$, and thus the convergence of the Dirichlet series is established.

4. Absolute Convergence of the Modified Exponential Series

THEOREM 5. *If the modified exponential series* (1.2) *is absolutely convergent for* $s_1 = \sigma_1 + it_1$ *where* $\sigma_1 > 0$, *then it is absolutely convergent in the half plane* $\Re(s) \geqq \sigma_1$.

PROOF. We need only show that for s in the half plane $\Re(s) \geqq \sigma_1$, given $\delta > 0$, we can find an integer n_0 such that for $q \geqq p \geqq n_0$ we have

$$\sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma) |G(\lambda_n s)| < \delta.$$

But by hypothesis, given $\epsilon > 0$, there exists an integer n_1 such that for $q \geqq p \geqq n_1$ we have

$$\sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma_1) |G(\lambda_n s_1)| < \epsilon.$$

We take $n_0 = \max [n_1, N_0(\sigma_1)]$ where N_0 is given by lemma 2. Then for $q \geqq p \geqq n_0$ we get

$$\sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma) |G(\lambda_n s)|$$

$$\leqq \sum_{n=p}^{n=q} |a_n| \exp(-\lambda_n \sigma_1) |G(\lambda_n s_1)| \left| \frac{G(\lambda_n s)}{G(\lambda_n s_1)} \right| \exp[(\sigma_1 - \sigma)\lambda_n]$$

$$\leqq \epsilon K_4(\sigma_1)$$

by use of the above and of lemma 4, and thus the proof is completed.

Similar to the case of convergence, there may exist a half plane of absolute convergence for the modified exponential series, and β will be used to denote the *abscissa of absolute convergence* and the equation $\Re(s) = \beta$ will be used to denote the *line of absolute convergence.*

The following two theorems will show equi-absolute convergence of the Dirichlet series and the modified exponential series.

THEOREM 6. *If the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_1)$ *is absolutely convergent and* $\Re(s_1) = \sigma_1 > 0$, *then the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_1) G(\lambda_n s_1)$ *is also absolutely convergent.*

PROOF. The proof is essentially a repetition of the argument of theorem 3 except that lemma 7 must be used instead of lemma 6.

THEOREM 7. *If the modified exponential series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is absolutely convergent for* $s = s_1 = \sigma_1 + it_1$, *where* $\sigma_1 > 0$, *then the Dirichlet series* $\sum_{n=1}^{\infty} a_n \exp(-\lambda_n s_1)$ *is also absolutely convergent.*

PROOF. The proof follows that of theorem 4 except that lemma 7 must be used instead of lemma 6.

5. Uniform Convergence of the Modified Exponential Series

THEOREM 8. *If the modified exponential series* $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is absolutely convergent for* $s = s_1$, $\Re(s_1) = \beta > 0$, *then the series is uniformly convergent for* $\Re(s) \geqq \beta$.

PROOF. The method of proof for theorem 5 establishes also the above theorem.

However, it may happen that the series is uniformly convergent for a certain range of values of s for which it is not absolutely convergent. Indeed, the series will in general have a half plane of uniform convergence, and we use γ to denote the *abscissa of uniform convergence.* It follows at once that the three abscissae of convergence satisfy the inequalities $\alpha \leqq \gamma \leqq \beta$.

Furthermore, the Dirichlet series and the modified exponential series are equi-uniformly convergent as is shown by the following two theorems.

THEOREM 9. *If the Dirichlet series* $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ *is uniformly convergent for* $\Re(s) \geqq \gamma > 0$, *then the modified exponential series* $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is also uniformly convergent for* $\Re(s) \geqq \gamma > 0$.

PROOF. In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s)$$

$$V_n = G(\lambda_n s).$$

Now, by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for $\Re(s) \geqq \gamma$ and for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s) \right| < \epsilon.$$

Thus for $q \geqq p \geqq n_0$ and for $\Re(s) \geqq \gamma$ we get

$$\begin{aligned}\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| &\leqq \left| \sum_{n=p}^{n=q-1} (G(\lambda_n s) - G(\lambda_{n+1} s)) \sum_{m=p}^{m=n} a_m \exp(-\lambda_m s) \right| \\ &\quad + |G(\lambda_q s)| \left| \sum_{m=p}^{m=q} a_m \exp(-\lambda_m s) \right| \\ &\leqq \epsilon \left[\sum_{n=p}^{n=q-1} |G(\lambda_n s) - G(\lambda_{n+1} s)| + |G(\lambda_q s)| \right].\end{aligned}$$

Now choose $n_1 = \max [n_0, N_2(\gamma)]$ where N_2 is given by lemma 6. Then by use of lemmas 1 and 6, for $q \geqq p \geqq n_1$ and for all s satisfying $\Re(s) \geqq \gamma$ we get

$$\begin{aligned}\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| &\leqq \epsilon \left[\sum_{n=p}^{n=q-1} \left| \int_{\lambda_n}^{\lambda_{n+1}} -\frac{\partial G(xs)}{\partial x} dx \right| + |G(\lambda_q s)| \right] \\ &\leqq \epsilon \left[\int_{\lambda_p}^{\lambda_q} \left| \frac{\partial G(xs)}{\partial x} \right| dx + |G(\lambda_q s)| \right] \\ &\leqq \epsilon \left[\frac{K_8(\gamma)}{|s|\lambda_p} + K_1(\gamma) \right] \leqq \epsilon \left[\frac{K_8(\gamma)}{\gamma \lambda_1} + K_1(\gamma) \right] \\ &\leqq \epsilon K(\gamma)\end{aligned}$$

thereby completing the proof.

THEOREM 10. *If the modified exponential series* $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ *is uniformly convergent for* $\Re(s) \geqq \gamma > 0$, *then the Dirichlet series* $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ *is uniformly convergent for* $\Re(s) \geqq \gamma > 0$.

PROOF. In Abel's lemma we identify

$$U_n = a_n \exp(-\lambda_n s) G(\lambda_n s),$$

$$V_n = \frac{1}{G(\lambda_n s)}.$$

Now by hypothesis, given $\epsilon > 0$, there exists an integer n_0 such that for all s satisfying $\Re(s) \geqq \gamma > 0$ and for $q \geqq p \geqq n_0$ we have

$$\left| \sum_{n=p}^{n=q} a_n \exp(-\lambda_n s) G(\lambda_n s) \right| < \epsilon.$$

Therefore for $q \geqq p \geqq n_0$ and for all s satisfying $\Re(s) \geqq \gamma > 0$ we get

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s)\right| \leqq \left|\sum_{n=p}^{n=q-1}\left(\frac{1}{G(\lambda_n s)} - \frac{1}{G(\lambda_{n+1} s)}\right)\sum_{m=p}^{m=n} a_m \exp(-\lambda_m s)G(\lambda_m s)\right| + \left|\frac{1}{G(\lambda_q s)}\right|\left|\sum_{m=p}^{m=q} a_m \exp(-\lambda_m s)G(\lambda_m s)\right|$$

$$\leqq \epsilon\left[\sum_{n=p}^{n=q-1}\left|\frac{1}{G(\lambda_n s)} - \frac{1}{G(\lambda_{n+1} s)}\right| + \left|\frac{1}{G(\lambda_q s)}\right|\right].$$

Now choose $n_1 = \max[n_0, N_0(\gamma), N_2(\gamma)]$ where N_0 and N_2 are given by lemmas 2 and 6. Then, by the use of these lemmas, for $q \geqq p \geqq n_1$, and for all s satisfying $\Re(s) \geqq \gamma$ we get

$$\left|\sum_{n=p}^{n=q} a_n \exp(-\lambda_n s)\right| \leqq \left[\sum_{n=p}^{n=q-1}\left|\int_{\lambda_n}^{\lambda_{n+1}} \frac{\frac{\partial G(xs)}{\partial x}}{[G(xs)]^2}\,dx\right| + \left|\frac{1}{G(\lambda_q s)}\right|\right]$$

$$\leqq \epsilon\left[\frac{K_s(\gamma)}{|s|\,[K_2(\gamma)]^2\lambda_p} + \frac{1}{K_2(\gamma)}\right] \leqq \epsilon\left[\frac{K_s(\gamma)}{\gamma\lambda_1[K_2(\gamma)]^2} + \frac{1}{K_2(\gamma)}\right]$$

$$\leqq \epsilon K(\gamma),$$

thereby completing the proof.

6. Consequences of Convergence

As a result of theorem 2 a number of properties of the modified exponential series are readily deduced. We state a few of these.

THEOREM 11. *If D is any finite region in the s plane such that for all points of D we have $\sigma \geqq \alpha + \delta > \alpha > 0$, then the modified exponential series (1.2) is uniformly convergent throughout $\bar{D}$, and its sum, the function $g(s)$, represents an analytic function in D.*

PROOF. This is an immediate consequence of theorem 2 and Weierstrass's classical theorem about uniformly convergent series of analytic functions.[7]

COROLLARY. *The derivative of $g(s)$ may be computed by termwise differentiation.*

THEOREM 12. *If the modified exponential series is convergent for the point $s_0 = \sigma_0 + it_0$ (where $\sigma_0 > 0$) to the value $g(s_0)$, then $g(s) \to g(s_0)$ whenever $s \to s_0$ along any path which lies in the angular sector $|\operatorname{am}(s - s_0)| \leqq \psi < \pi/2$.*

PROOF. This is an immediate consequence of theorem 2.

THEOREM 13. *If the modified exponential series $g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)G(\lambda_n s)$ is convergent for $s = s_0 = \sigma_0 + it_0$ where $\sigma_0 > 0$, and if E denotes the angular sector*

$$|\operatorname{am}(s - s_0)| \leqq \psi < \frac{\pi}{2}$$

[7] Titchmarsh, "The Theory of Functions," page 95.

and if $g(s) = 0$ for an infinity of values of s lying in E, then $a_n = 0$ for all values of n.

PROOF. We assume $g(s) \not\equiv 0$, else the proof is trivial. Since $g(s)$ is analytic in E, in any finite neighborhood of a point of E there must be only a finite

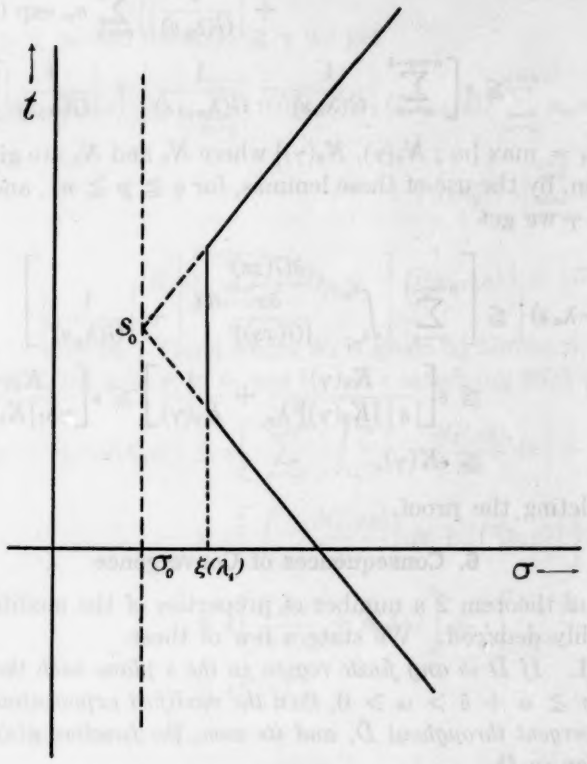


FIG. 3

number of zeros of $g(s)$. Because of the angular character of the set E it must be possible to select a sequence of values $s_n = \sigma_n + it_n$ where

(a) $\sigma_n > \xi(\lambda_1)$, ξ given by lemma 3

(b) $\lim_{n\to\infty} \sigma_n = \infty$

and such that $g(s_n) = 0$ for each n.

We consider the function

$$r(s) = \frac{\exp(\lambda_1 s)}{G(\lambda_1 s)}\, g(s)$$

$$= a_1 + \sum_{n=2}^{\infty} a_n \exp[-(\lambda_n - \lambda_1)s]\, \frac{G(\lambda_n s)}{G(\lambda_1 s)}.$$

This new function $r(s)$ is convergent for $s = s_0$ and is uniformly convergent in E' where E' denotes the deleted (if $\xi(\lambda_1) > \sigma_0$) angular sector

$$\sigma \geqq \max [\xi(\lambda_1), \sigma_0] > 0$$

$$|\operatorname{am} (s - s_0)| \leqq \psi < \frac{\pi}{2}.$$

We let $s \to \infty$ along any set of values whatsoever in E'. Since for a uniformly convergent series we can take the termwise limit,[8] we get $r(s) \to a_1$. But if we let $s \to \infty$ over the set $\{s_n\}$ we get $r(s) \to 0$. Thus $a_1 = 0$. Since this argument may be repeated for each n, the theorem follows at once.[9]

THEOREM 14. *If the two series*

$$\psi(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$$

$$\varphi(s) = \sum_{n=1}^{\infty} b_n \exp(-\lambda_n s) G(\lambda_n s)$$

are convergent at $s = s_0 = \sigma_0 + it_0$ where $\sigma_0 > 0$ and if in an angular sector E there are an infinity of values s_n such that $\psi(s_n) = \varphi(s_n)$ then $a_n = b_n$ for all n.

PROOF. On writing $g(s) = \psi(s) - \varphi(s)$, we see that the previous theorem applies, and thus we get our conclusion.

As a consequence of the sets of equi-convergence theorems, we have the following results:[10]

THEOREM 15. *If the abscissa of convergence α of the modified exponential series*

$$g(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$$

is positive, then it is given by the formula

$$\alpha = \overline{\lim_{n \to \infty}} \frac{\log \left| \sum_{m=1}^{n} a_m \right|}{\lambda_n}.$$

THEOREM 16. *If the abscissa of absolute convergence β of the modified exponential series is positive, then it is given by*

$$\beta = \overline{\lim_{n \to \infty}} \frac{\log \sum_{m=1}^{n} |a_m|}{\lambda_n}.$$

---

[8] Titchmarsh, l.c., page 8.

[9] This proof is essentially borrowed from Hardy and Riesz, l.c., page 6.

[10] Hardy and Riesz, l.c., pages 7–9.

THEOREM 17. *We have*

$$\beta - \alpha \leqq \overline{\lim_{n\to\infty}} \frac{\log n}{\lambda_n}.$$

7. Convergence on the Line of Convergence for the Hankel Series

For Dirichlet series it is well known that under certain special conditions the real point of the line of convergence is a singularity of the function.[11] The following analogous theorem holds for the Hankel series (1.9).

THEOREM 18. *If ν is real, $\nu \geqq \frac{1}{2}$, and if all the coefficients of the Hankel series are positive or zero, and if $\Re(s) = \alpha > 0$ defines the line of convergence, then the point $s = \alpha$ is a singular point of the function represented by the series.*

PROOF. Let the series be represented by

$$h(s) = \sum_{n=1}^{\infty} a_n (i\lambda_n s)^{\frac{1}{2}} H_\nu^{(1)}(i\lambda_n s)$$

$$= \sum_{n=1}^{\infty} a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s).$$

It will be convenient to use the function $K_\nu(z)$ defined by[12]

$$K_\nu(z) = \frac{i\pi}{2} \exp\left(\frac{\nu\pi i}{2}\right) H_\nu^{(1)}(iz). \tag{7.1}$$

This function has the integral representation[13]

$$K_\nu(z) = \int_0^\infty \exp(-z \cosh t) \cosh \nu t \, dt. \tag{7.2}$$

Thus we consider the function

$$h(s) = (i)^{\frac{1}{2}} \frac{2}{\pi i} \exp\left(-\frac{\nu\pi i}{2}\right) \sum_{n=1}^{\infty} a_n (\lambda_n s)^{\frac{1}{2}} K_\nu(\lambda_n s) \tag{7.3}$$

which is analytic at the point $s = \alpha + 1 = \alpha_1$, and hence can be expanded in a Taylor's series about this point. We shall assume that the point $s = \alpha$ is not a singular point of the function $h(s)$, then the above Taylor's series expansion must be valid for a value of s smaller than α, say $\alpha_2 < \alpha$. By justifying certain summation interchanges we shall show that this requires $h(s)$ to be convergent at the point $s = \alpha_2 < \alpha$, thereby establishing a contradiction.

The domain in which we are interested does not contain the origin, so that the function

$$r(s) = \frac{\pi i}{2(i)^{\frac{1}{2}}} \exp\left(\frac{\nu\pi i}{2}\right) \frac{h(s)}{(s)^{\frac{1}{2}}} \tag{7.4}$$

---

[11] Hardy and Riesz, l.c., page 10.

[12] Watson, l.c., page 78.

[13] Watson, l.c., page 181.

will have the same singularities as $h(s)$. The Taylor's series expansion of the function $r(s)$ about the point $s = \alpha_1$ will, by hypothesis, converge at the point $s = \alpha_2$. Indeed

$$r(s) = \frac{\pi i}{2(i)^{\frac{1}{2}}} \exp\left(\frac{\nu\pi i}{2}\right) \frac{h(s)}{(s)^{\frac{1}{2}}}$$

$$= \sum_{n=1}^{\infty} a_n(\lambda_n)^{\frac{1}{2}} \int_0^{\infty} \exp(-\lambda_n s \cosh t) \cosh \nu t \, dt$$

and

$$r(s) = \sum_{m=0}^{\infty} \frac{(s-\alpha_1)^m}{m!} r^{(m)}(\alpha_1)$$

$$= \sum_{m=0}^{\infty} \frac{(s-\alpha_1)^m}{m!} \sum_{n=1}^{\infty} a_n(\lambda_n)^{\frac{1}{2}}(-1)^m(\lambda_n)^m \cdot \int_0^{\infty} \exp(-\alpha_1\lambda_n \cosh t)(\cosh t)^m \cosh \nu t \, dt.$$

We note that the value of the integral is positive since the integrand is always positive, that $(-1)^m(\alpha_2 - \alpha_1)^m$ is positive, and we recall that we assumed the coefficients a_n were all positive. Hence we can interchange the order of summation[14] and we have

$$r(\alpha_2) = \sum_{n=1}^{\infty} a_n(\lambda_n)^{\frac{1}{2}} \sum_{m=0}^{\infty} \frac{(\alpha_1 - \alpha_2)^m}{m!} \int_0^{\infty} \exp(-\alpha_1\lambda_n \cosh t)(\lambda_n \cosh t)^m \cosh \nu t \, dt.$$

We recognize that

$$\sum_{m=0}^{\infty} \frac{(\alpha_1 - \alpha_2)^m}{m!} \int_0^{\infty} \exp(-\alpha_1\lambda_n \cosh t)(\lambda_n \cosh t)^m \cosh \nu t \, dt$$

is the Taylor's series expansion for $\int_0^{\infty} \exp(-\alpha_2\lambda_n \cosh t) \cosh \nu t \, dt$ about the point α_1.

Thus we have that $r(\alpha_2)$ converges,

$$r(\alpha_2) = \sum_{n=1}^{\infty} a_n\sqrt{\lambda_n} \int_0^{\infty} \exp(-\alpha_2\lambda_n \cosh t) \cosh \nu t \, dt$$

$$= \sum_{n=1}^{\infty} a_n\sqrt{\lambda_n} K_\nu(\lambda_n\alpha_2).$$

Therefore the series

$$h(\alpha_2) = \frac{2\sqrt{i}}{\pi i} \exp\left(\frac{-\nu\pi i}{2}\right) \sqrt{\alpha_2}\, r(\alpha_2) = \sum_{n=1}^{\infty} a_n\sqrt{i\lambda_n s}\, H_\nu^{(1)}(i\lambda_n s)$$

$$= \sum_{n=1}^{\infty} a_n A_\nu \exp(-\lambda_n s) M_\nu(\lambda_n s)$$

[14] Bromwich, "Infinite Series," page 78.

converges, and since $\alpha_2 < \alpha$ we have a contradiction, α being defined as the abscissa of convergence. Hence the value $s = \alpha$ must be a singular point for the function $h(s)$.

8. Inversion Formulae

For the Dirichlet series $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ Perron has given the following formula for the sum of the coefficients

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{f(s)}{s} \exp(\omega s)\, ds = \sum_{m=1}^{n} a_m, \tag{8.1}$$

where

$$\lambda_n < \omega < \lambda_{n+1},$$
$$c > \alpha.$$

In order to establish related formulae for the modified exponential series, we shall first prove a theorem concerning the order of the function $g(s)$ as $t \to \infty$.

THEOREM 19. *If the series*

$$g(s) = \sum_{m=1}^{\infty} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s),$$

where s is the complex variable $s = \sigma + it$ and where $0 \leqq h \leqq \lambda_1$, is convergent or finitely oscillating at the point $s = \sigma_0 + 0$, then given $\epsilon > 0$, $\delta > 0$ we can find a value t_0 such that for $|t| \geqq t_0$, $\sigma \geqq \sigma_0 + \epsilon$, and for all n we have

$$\left|\frac{1}{t}\sum_{1}^{n} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)\right| < \delta, \tag{8.2}$$

or

$$\sum_{1}^{n} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = o(|t|).^{15} \tag{8.3}$$

In particular, for $n = \infty$, we have

$$g(s) = o(|t|) \tag{8.4}$$

which holds uniformly for $\sigma \geqq \sigma_0 + \epsilon$.

PROOF. Because of the convergence or finite oscillation of the series at $s = \sigma_0$, and because of the fact that there are only a finite number of terms of type (a) below, there exists a constant L such that for all integers m, p, and $q \geqq p$, and for $r \leqq \max[N_0(\sigma_0), N_1(\sigma_0)]$ where N_0 and N_1 are given by lemmas 2 and 5 we obtain

$$|a_r \exp[-(\lambda_r - h)\sigma_0]| < L \tag{a}$$

[15] The notation $F(x) = o(\varphi(x))$ means that as x approaches a definite limit, $\lim \frac{F(x)}{\varphi(x)} = 0$.

(b) $$|a_m \exp[-(\lambda_m - h)\sigma_0]G(\lambda_m\sigma_0)| < L$$

(c) $$\left|\sum_{m=p}^{q} a_m \exp[-(\lambda_m - h)\sigma_0]G(\lambda_m\sigma_0)\right| < L.$$

Let N be a function of $|t|$, say $N(|t|)$. We consider sums of the form

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)$$

and treat two cases, depending on whether $N < n$ or $N \geqq n$.

CASE 1. Let

(8.5) $$1 \leqq \max[N_0, N_1] < N < n.$$

We write

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = \sum_1^{N-1} + \sum_N^n.$$

Apply Abel's lemma to the summation $\sum_N^n$ to get

$$\begin{aligned}\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)&\\ = \sum_1^{N-1} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) &+ \sum_{m=N}^{n-1}\left[\exp[-(s-\sigma_0)(\lambda_m - h)]\frac{G(\lambda_m s)}{G(\lambda_m\sigma_0)}\right.\\ &\left. - \exp[-(s-\sigma_0)(\lambda_{m+1} - h)]\frac{G(\lambda_{m+1}s)}{G(\lambda_{m+1}\sigma_0)}\right]\\ &\cdot\left[\sum_{p=N}^{p=m} a_p \exp[-(\lambda_p - h)\sigma_0]G(\lambda_p\sigma_0)\right]\\ + \left[\sum_{m=N}^{m=n} a_m \exp[-(\lambda_m - h)\sigma_0]G(\lambda_m\sigma_0)\right]&\exp[-(s-\sigma_0)(\lambda_n - h)]\frac{G(\lambda_n s)}{G(\lambda_n\sigma_0)}\end{aligned} \tag{8.6}$$

$$= S_1 + S_2 + S_3.$$

We have

$$\begin{aligned}|S_1| &= \left|\sum_1^{N-1} a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)\right|\\ &\leqq \sum_1^{N-1} |a_m \exp[-(\lambda_m - h)\sigma_0]|\cdot K_1(\sigma_0)\\ &\leqq (N-1)\cdot L\cdot K_1\end{aligned} \tag{8.7}$$

by use of lemma 1 and inequality (a). Also

$$|S_3| = \left|\sum_{m=N}^{n} a_m \exp[-(\lambda_m - h)\sigma_0)]G(\lambda_m \sigma_0)\right| \cdot \left|\exp[-(s-\sigma_0)(\lambda_n - h)]\frac{G(\lambda_n s)}{G(\lambda_n \sigma_0)}\right| \tag{8.8}$$

$$\leqq L \cdot K_4(\sigma_0)$$

$$\leqq L \cdot K_4$$

by use of lemma 4 and inequality (c).

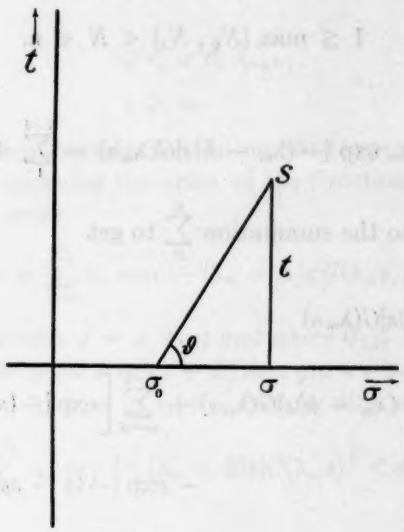


FIG. 4

Now the sum

$$S_2 = \sum_{m=N}^{n-1}\left[\exp[-(\lambda_m - h)(s-\sigma_0)]\frac{G(\lambda_m s)}{G(\lambda_m \sigma_0)} - \exp[-(\lambda_{m+1} - h)(s-\sigma_0)]\frac{G(\lambda_{m+1} s)}{G(\lambda_{m+1}\sigma_0)}\right]\left[\sum_{p=N}^{m} a_p \exp[-(\lambda_p - h)\sigma_0]G(\lambda_p \sigma_0)\right]$$

save for the factor $\exp[h(s-\sigma_0)]$ can be dealt with in the same way as was done in the proof of theorem 2, and similar to relation (3.4) we get

$$|S_2| \leqq L \cdot K_{10}(\sigma_0)\frac{|s-\sigma_0|}{\Re(s-\sigma_0)}\exp[-(\lambda_N - h)\Re(s-\sigma_0)].$$

Now suppose am $(s-\sigma_0) = \theta$. Then

$$\frac{|s-\sigma_0|}{\Re(s-\sigma_0)} = \sec\theta = \sqrt{1+\tan^2\theta} \leqq \sqrt{1+\frac{t^2}{\epsilon^2}}$$

for $\sigma \geqq \sigma_0 + \epsilon$. Thus we have

$$|S_2| \leqq L \cdot K_{10} \cdot \sqrt{1 + \frac{t^2}{\epsilon^2}} \exp[-(\lambda_N - h)\epsilon]$$

or

$$S_2 = O(|t \exp[-(\lambda_N - h)\epsilon]|).^{16} \tag{8.9}$$

Therefore, substituting (8.7), (8.8), and (8.9) in relation (8.6) we get for $1 \leqq \max[N_0, N_1] < N < n$

$$\left|\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)\right| = |S_1 + S_2 + S_3|$$

$$\leqq (N-1) \cdot L \cdot K_1 + L \cdot K_{10}\sqrt{1 + \frac{t^2}{\epsilon^2}} \exp[-(\lambda_N - h)\epsilon] + L \cdot K_4$$

or

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = O(N) + O(|t \exp[-(\lambda_N - h)\epsilon]|). \tag{8.10}$$

CASE 2. For $n \leqq N$ we get at once

$$\left|\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s)\right| \leqq nL \cdot K_1 \leqq N \cdot L \cdot K_1$$

or

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = O(N). \tag{8.11}$$

Now let $N(|t|)$ be so chosen that $N(|t|)$ tends to infinity more slowly than $|t|$, and also so that relation (8.5) is satisfied. Then for either of the two cases for $\Re(s) = \sigma \geqq \sigma_0 + \epsilon$ and for all n we have

$$\sum_1^n a_m \exp[-(\lambda_m - h)s]G(\lambda_m s) = o(|t|), \tag{8.12}$$

thereby completing the proof.

LEMMA 10. *If $b > 0$, $k \geqq 0$, then*

$$\frac{1}{2\pi i}\int_{b-i\infty}^{b+i\infty} \exp(\omega s)\frac{ds}{s^{k+1}} = \frac{\omega^k}{\Gamma(k+1)}, \qquad \omega \geqq 0,$$

$$= 0, \qquad \omega < 0.$$

See Hardy and Riesz for proof.[17]

[16] The notation $F(x) = O(\varphi(x))$ means that $|F(x)| < K\varphi(x)$ for values of x sufficiently near a given limiting value.

[17] Hardy and Riesz, l.c., page 50.

THEOREM 20. *(First Inversion Formula) If* $\lambda_n < \omega < \lambda_{n+1}$ *and if the modified exponential series* $g(s) = \sum_1^\infty a_m \exp(-\lambda_m s) G(\lambda_m s)$ *is convergent for* $s = \sigma_0 + it_0$ *where* $\sigma_0 > \xi(\lambda_1)$ *(*ξ *being given by lemma 3), then for* $c > \sigma_0$ *we have*

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{g(s)}{s}\frac{\exp(\omega s)}{G(\lambda_n s)}\,ds = a_n + \sum_1^{n-1} \frac{a_m}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp[-(\lambda_m - \omega)s]\frac{G(\lambda_m s)}{G(\lambda_n s)}\frac{ds}{s}.$$

PROOF. We consider the function

$$r(s) = \frac{\exp[\omega(s-\sigma_0)]}{G(\lambda_n s)}\left\{g(s) - \sum_1^n a_m \exp(-\lambda_m s)G(\lambda_m s)\right\} \tag{8.13}$$

$$= \sum_{m=n+1}^{\infty} a_m \exp[-(\lambda_m - \omega)s]\frac{G(\lambda_m s)}{G(\lambda_n s)}\exp(-\omega\sigma_0)$$

where n is determined so that $\lambda_n < \omega < \lambda_{n+1}$.

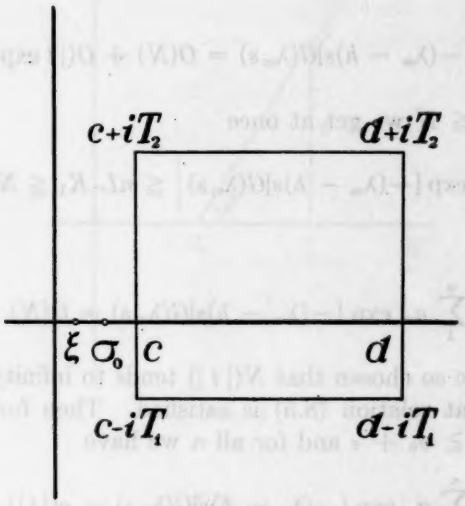


FIG. 5

We shall establish that

$$\int_{c-i\infty}^{c+i\infty} \frac{r(s)}{s}\,ds = 0.$$

We consider the closed rectangle as shown at the side. The function $r(s)/s$ has no singularities in this rectangle, and so by Cauchy's theorem we have

$$\int_{c-iT_1}^{c+iT_2} \frac{r(s)}{s}\,ds = \int_{c-iT_1}^{d-iT_1} \frac{r(s)}{s}\,ds - \int_{c+iT_2}^{d+iT_2} \frac{r(s)}{s}\,ds + \int_{d-iT_1}^{d+iT_2} \frac{r(s)}{s}\,ds. \tag{8.14}$$

Now this finite rectangle can be enclosed in an angular sector about the point $s_0 = \sigma_0 + it_0$, and by using the same argument as we did in theorem 2, given

$\epsilon > 0$, there exists an integer n_0 such that $n_0 \geqq n$ and for which for $p \geqq n_0$ we have

$$\left|\sum_{m=p}^{\infty} a_m \exp\left[-(\lambda_m - \omega)s\right] \frac{G(\lambda_m s)}{G(\lambda_n s)} \exp\left(-\omega\sigma_0\right)\right|$$

$$\leqq \frac{\epsilon K_{10}(\sigma_0)}{K_3(\xi)} \frac{|s - \sigma_0|}{\Re(s - \sigma_0)} \exp\left[-\Re(s - \sigma_0)\lambda_p\right] \exp\left(-\omega\sigma_0\right) \exp\left(\Re(s)\omega\right)$$

$$\leqq \frac{\epsilon K_{10}}{K_3} \frac{|s - \sigma_0|}{\Re(s - \sigma_0)} \exp\left[-(\lambda_p - \omega)\Re(s - \sigma_0)\right] \tag{8.15}$$

$$\leqq \epsilon K_{11},$$

this estimate being independent of s as long as s is in the angular sector, and hence independent of the number d. Thus it follows that there exists a constant K_{12} such that

$$\left|\sum_{m=n+1}^{\infty} a_m \exp\left[-(\lambda_m - \omega)s\right] \frac{G(\lambda_m s)}{G(\lambda_n s)} \exp\left(-\omega\sigma_0\right)\right| < K_{12},$$

for all s in the angular sector.

Therefore

$$\left|\int_{d-iT_1}^{d+iT_2} \frac{r(s)}{s}\, ds\right| \leqq K_{12}\left|\int_{d-iT_1}^{d+iT_2} \frac{ds}{s}\right| = K_{12}\left|\log\frac{d + iT_2}{d - iT_1}\right|$$

$$\leqq K_{12}\left|\log\frac{1 + \dfrac{iT_2}{d}}{1 - \dfrac{iT_1}{d}}\right|, \tag{8.16}$$

and as $d \to \infty$ we have

$$\left|\log\frac{1 + \dfrac{iT_2}{d}}{1 - \dfrac{iT_1}{d}}\right| \to 0. \tag{8.17}$$

We shall now show that

$$\int_{c-iT_1}^{d-iT_1} \frac{r(s)}{s}\, ds \quad \text{and} \quad \int_{c+iT_2}^{d+Ti_2} \frac{r(s)}{s}\, ds$$

are both convergent as $d \to \infty$, and thus we can write (8.14) as

$$\int_{c-iT_1}^{c+iT_2} \frac{r(s)}{s}\, ds = \int_{c-iT_1}^{\infty-iT_1} \frac{r(s)}{s}\, ds - \int_{c+iT_2}^{\infty+iT_2} \frac{r(s)}{s}\, ds. \tag{8.18}$$

If we let the equation

$$r(s) = \frac{\exp\left[-(\lambda_{n+1} - \omega)s\right]}{G(\lambda_n s)} p(s) \tag{8.19}$$

define the function $p(s)$ we have

$$\begin{aligned} p(s) &= r(s) \exp [(\lambda_{n+1} - \omega)s] G(\lambda_n s) \\ &= \left[\sum_{m=n+1}^{\infty} a_m \exp [-(\lambda_m - \lambda_{n+1})s] G(\lambda_m s) \exp (-\omega\sigma_0) \right]. \end{aligned} \tag{8.20}$$

Now $r(s)$ is convergent at the point $s = \sigma_0$. Therefore the series for the function $p(s)$ converges for $s = \sigma_0$. Identifying

$$b_p = a_{n+p},$$

$$\lambda_{n+p} = \mu_p,$$

we see that theorem 19 is applicable to the function

$$p(s) = \sum_{p=1}^{\infty} b_p \exp [-(\mu_p - \mu_1)s] G(\mu_p s) \exp (-\omega\sigma_0).$$

Therefore, given $\epsilon_1 > 0$, we can choose T such that $T > 0$ and

$$|p(s)| < \epsilon_1 T_2 \quad \text{for} \begin{cases} s = \sigma + iT_2 \\ \sigma \geqq c \\ T_2 > T \end{cases}.$$

Then

$$\begin{aligned} \left| \int_{c+iT_2}^{d+iT_2} \frac{r(s)}{s} ds \right| &= \left| \int_{c+iT_2}^{d+iT_2} \frac{p(s)}{s} \frac{\exp [-(\lambda_{n+1} - \omega)s]}{G(\lambda_m s)} ds \right| \\ &\leqq \epsilon_1 T_2 \left| \int_{c+iT_2}^{d+iT_2} \frac{\exp [-(\lambda_{n+1} - \omega)s]}{sG(\lambda_m s)} ds \right| \\ &\leqq \frac{\epsilon_1 T_2}{K_3(\xi)} \frac{1}{\sqrt{c^2 + T_2^2}} \int_c^{\infty} \exp [-(\lambda_{n+1} - \omega)s] ds \\ &< \frac{\epsilon_1}{K_3} \frac{1}{\lambda_{n+1} - \omega}, \end{aligned} \tag{8.21}$$

since $\frac{1}{|s|} \leqq \frac{1}{\sqrt{c^2 + T_2^2}}$ and $\frac{T_2}{\sqrt{c^2 + T_2^2}} < 1$. Thus we conclude that

$$\lim_{T_2 \to \infty} \left| \int_{c+iT_2}^{\infty+iT_2} \frac{r(s)}{s} ds \right| = 0, \tag{8.22}$$

and likewise

$$\lim_{T_1 \to \infty} \left| \int_{c-iT_1}^{\infty-iT_1} \frac{r(s)}{s} ds \right| = 0. \tag{8.23}$$

Therefore from (8.19) and the above we obtain the evaluation

$$\int_{c-i\infty}^{c+i\infty} \frac{r(s)}{s} ds = 0,$$

and recalling the definition of $r(s)$ (equation (8.13)) we have

$$\int_{c-i\infty}^{c+i\infty} \frac{g(s)}{s} \frac{\exp[\omega(s-\sigma_0)]}{G(\lambda_n s)}\,ds = \int_{c-i\infty}^{c+i\infty} \left[\sum_{m=1}^{n} a_m \exp[-\lambda_m s] \frac{G(\lambda_m s)}{G(\lambda_n s)}\right] \exp[\omega(s-\sigma_0)] \frac{ds}{s}.$$

Using lemma 10 to evaluate the term in which a_n is a factor we obtain the first inversion formula,

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{g(s)}{s} \frac{\exp(\omega s)}{G(\lambda_n s)}\,ds = a_n + \sum_{1}^{n-1} \frac{a_m}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp[-(\lambda_m - \omega)s] \frac{G(\lambda_m s)}{G(\lambda_n s)} \frac{ds}{s}.$$

THEOREM 21. *(Second Inversion Formula) If the series*

$$g(s) = \sum_{m=1}^{\infty} a_m \exp(-\lambda_m s) G(\lambda_m s)$$

is convergent for $s = \sigma_0 + it_0$ *and if* $\lambda_n < \omega < \lambda_{n+1}$, *then for* $c \geqq \sigma_0$ *we have*

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{g(s)\exp(\omega s)}{s}\,ds = \frac{1}{2\pi i}\sum_{1}^{n} a_m \int_{c-i\infty}^{c+i\infty} \exp[-(\lambda_m - \omega)s] G(\lambda_m s) \frac{ds}{s}.$$

PROOF. The proof follows along the same lines as did the proof of theorem 20 except that we do not have to introduce the factor $1/G(\lambda_n s)$.

REMARKS: For the Hankel series and for the special cases when ν is an odd half-integer, the second inversion formula can be simplified. Let $\nu - \frac{1}{2} = k$ where k is an integer. It will be recalled that the Bessel functions have simple expressions for such values, and likewise the function $M_\nu(\lambda_m s)$ can be expressed by the finite series

$$\begin{aligned} M_\nu(\lambda_m s) &= \int_0^\infty \exp(-u) u^k \left(1 + \frac{u}{2\lambda_m s}\right)^k du \\ &= \int_0^\infty \exp(-u) u^k \left[1 + \binom{k}{1}\frac{u}{2\lambda_m s} + \binom{k}{2}\left(\frac{u}{2\lambda_m s}\right)^2 + \cdots + \left(\frac{u}{2\lambda_m s}\right)^k\right] du \\ &= k! + \binom{k}{1}\frac{(k+1)!}{2\lambda_m s} + \binom{k}{2}\frac{(k+2)!}{(2\lambda_m s)^2} + \cdots + \frac{(2k)!}{(2\lambda_m s)^k}, \end{aligned}$$

and on substituting this in theorem 21 and using lemma 10 we get

$$\begin{aligned} \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{h(s)\exp(\omega s)}{s}\,ds = \sum_{1}^{n} a_m \Bigg[k! + \binom{k}{1}\frac{(k+1)!}{2\lambda_m}\frac{(w-\lambda_m)}{1!} \\ + \binom{k}{2}\frac{(k+2)!}{(2\lambda_m)^2}\frac{(w-\lambda_m)^2}{2!} + \cdots + \frac{(2k)!}{(2\lambda_m)^k}\frac{(w-\lambda_m)^k}{k!}\Bigg]. \end{aligned}$$

We recall that for $\nu = \frac{1}{2}$ the Hankel series reduces to the Dirichlet series, and we note that the above formula reduces to Perron's formula (8.1) for $\nu - \frac{1}{2} = k = 0$.

9. Conclusion

By defining

$$
\begin{aligned}
A(x) &= 0, \qquad 0 \leqq x < \lambda_1, \\
A(\lambda_1) &= \frac{a_1}{2}, \\
A(\lambda_n) &= \sum_{m=1}^{n-1} a_m + \frac{a_n}{2} \qquad (9.1) \\
A(x) &= \sum_{m=1}^{n} a_m, \qquad \lambda_n < x < \lambda_{n+1},
\end{aligned}
$$

we can express the Dirichlet series $f(s) = \sum_{n=1}^{\infty} a_n \exp(-\lambda_n s)$ as the Stieltjes integral $f(s) = \int_0^{\infty} \exp(-xs)\, d(A(x))$, and Perron's formula for the sum of the coefficients becomes

$$
A(x) = \frac{1}{2\pi i} \int_{-i\infty}^{c+i\infty} \frac{f(s)}{s} \exp(xs)\, ds. \qquad (9.2)
$$

Likewise the modified exponential series $g(s) = \sum_1^{\infty} a_n \exp(-\lambda_n s) G(\lambda_n s)$ becomes the Stieltjes integral

$$
g(s) = \int_0^{\infty} \exp(-xs) G(xs)\, d(A(x)). \qquad (9.3)
$$

It should be possible to obtain a convergence theorem similar to theorem 2 from a consideration of the Laplace-Stieltjes transform.[18]

An integration by parts of equation (9.3) would give formally

$$
g(s) = \int_0^{\infty} s \exp(-xs) \left\{ G(xs) + \frac{\partial G(xs)}{\partial x} \right\} A(x)\, dx, \qquad (9.4)
$$

which can be regarded as an integral equation with the kernel

$$
s \exp(-xs) \left\{ G(xs) + \frac{\partial G(xs)}{\partial x} \right\}, \qquad (9.5)
$$

and the two inversion formulae for the modified exponential series could be written in the form

$$
\int_0^{\infty} dA(x) \left[\int_{c-i\infty}^{c+i\infty} \exp[-(x-\omega)s] G(xs) \frac{ds}{s} \right] = \int_{c-i\infty}^{c+i\infty} \frac{g(s) \exp \omega s}{s}\, ds, \qquad (9.6)
$$

and

[18] Private communication from Prof. J. D. Tamarkin to the author.

$$(9.7)\quad \int_0^\infty dA(x)\left[\int_{c-i\infty}^{c+i\infty} \exp\left[-(x-\omega)s\right]\frac{G(xs)}{G(\lambda_n s)}\frac{ds}{s}\right] = \int_{c-i\infty}^{c+i\infty}\frac{g(s)\exp \omega s}{sG(\lambda_n s)}\,ds,$$

as can be formally verified by substituting for $g(s)$, and changing the order of integration.

Hadamard's results relative to the composition of singularities of Taylor's series show that the only possible singular points for a function represented by the series $\sum_{n=1}^{\infty} a_n b_n \exp(-ns)$ are points of the form $\alpha + \beta$ where α is some singular point of the function $f(s) = \sum_{n=1}^{\infty} a_n \exp(-ns)$ and β is some singular point of $g(s) = \sum_{n=1}^{\infty} b_n \exp(-ns)$. Mandelbrojt obtained analogous results applicable to the general Dirichlet series, but the present author has been unable to get an extension of Mandelbrojt's work applicable to the modified exponential series.

In conclusion I wish to thank Prof. S. Bochner of Princeton University who suggested this topic and whose advice and criticism have always been helpful.

PRINCETON UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 42, No. 3, July, 1941

ON THE INTEGRALS OF CANONICAL SYSTEMS

BY CARL LUDWIG SIEGEL

(Received December 19, 1940)

1. Trigonometrical series

We consider a *canonical system* of differential equations

$$\dot{x}_k = H_{y_k}, \qquad \dot{y}_k = -H_{x_k} \qquad (k = 1, \cdots, n) \tag{1}$$

and suppose that the real function H does not contain the independent variable t and is, in a neighbourhood of the origin, an *analytic* function of the $2n$ variables $x_1, \cdots, y_n$. Let the number n of degrees of freedom be at least 2. We suppose moreover that the origin is an *equilibrium point* of the system; i.e. all the $2n$ derivatives H_{x_k}, H_{y_k} $(k = 1, \cdots, n)$ vanish at the origin. It may also be assumed that the function H itself vanishes at the origin. Denoting the $2n$ variables $x_1, \cdots, y_n$ by $z_1, \cdots, z_{2n}$, we write the system (1) in the form

$$\dot{z}_k = \sum_{l=1}^{2n} a_{kl} z_l + R_k \qquad (k = 1, \cdots, 2n), \tag{2}$$

where R_k is a power series in $z_1, \cdots, z_{2n}$ beginning with terms of the second order.

Let $\lambda_1, \cdots, \lambda_{2n}$ be the characteristic roots of the matrix (a_{kl}). In our case of a canonical system, the characteristic polynomial is an even function; hence we may arrange the roots such that

$$\lambda_{k+n} = -\lambda_k \qquad (k = 1, \cdots, n). \tag{3}$$

We suppose that $\lambda_1, \cdots, \lambda_n$ are *linearly independent*, with respect to the field of rational numbers; this means that the relationship

$$g_1\lambda_1 + \cdots + g_n\lambda_n = 0$$

holds in integral numbers $g_1, \cdots, g_n$ only for $g_1 = 0, \cdots, g_n = 0$.

If w is a power series of $2n$ variables x_k, η_k $(k = 1, \cdots, n)$ without linear terms and the determinant $| w_{x_k\eta_l} | \neq 0$ at the origin, the equations

$$\xi_k = w_{\eta_k}, \qquad y_k = w_{x_k} \qquad (k = 1, \cdots, n) \tag{4}$$

define a *contact transformation* of the variables $x_1, \cdots, y_n$ into the variables $\xi_1, \cdots, \eta_n$, and the canonical system (1) is invariant:

$$\dot{\xi}_k = H_{\eta_k}, \qquad \dot{\eta}_k = -H_{\xi_k} \qquad (k = 1, \cdots, n). \tag{5}$$

It has been known[1] for a long time that after an appropriate contact transformation the function H will depend only upon the n products

$$\xi_k\eta_k = \zeta_k \qquad (k = 1, \cdots, n) \tag{6}$$

and take the form

$$H = \sum_{k=1}^{n} \lambda_k\zeta_k + R, \tag{7}$$

where R is a power series in the n variables $\zeta_1, \cdots, \zeta_n$ beginning with quadratic terms. The integration of (5) for this *normal form* of H is then immediate. Since

$$H_{\eta_k} = \xi_k H_{\zeta_k}, \qquad H_{\xi_k} = \eta_k H_{\zeta_k}, \tag{8}$$

we deduce from (5) that the n products ζ_k are constant; hence

$$\xi_k = \alpha_k e^{H_{\zeta_k}t}, \qquad \eta_k = \beta_k e^{-H_{\zeta_k}t} \qquad (k = 1, \cdots, n),$$

with arbitrary constants α_k, β_k and

$$\alpha_k\beta_k = \zeta_k \qquad (k = 1, \cdots, n).$$

The original unknown functions $x_1, \cdots, y_n$ become now series with the general term

$$c_{g_1\cdots g_n} e^{(g_1 H_{\zeta_1} + \cdots + g_n H_{\zeta_n})t},$$

where the coefficient $c_{g_1\cdots g_n}$ denotes a constant and $g_1, \cdots, g_n$ integers. If all the characteristic roots $\lambda_1, \cdots, \lambda_n$ are pure imaginary numbers and the initial values of $x_1, \cdots, y_n$ real, then all the values $H_{\zeta_1}, \cdots, H_{\zeta_n}$ are pure imaginary, and we get a representation of the solutions of the canonical system by trigonometrical series.

This elegant method of solution has also been generalized to the case of a function H which contains explicitly the variable t, in periodical form, and is closely related to the important researches of Delaunay, Hill and Poincaré[2] in celestial mechanics. However, there is a serious objection: *The question of convergence has never been settled.* If we define sum, difference, product, quotient and derivative of power series in a formal algebraic manner, we can perform these operations also with divergent power series, and then we can construct by straightforward calculation a transformation of the type (4) which reduces H to a power series of the n products $\xi_k\eta_k$ alone. But no proof for the convergence of this contact transformation has been given, with exception of some special examples, when the integration of the system (1) can also be carried out by elementary methods. On account of the small divisors appearing in the

[1] E. T. Whittaker, *On the solution of dynamical problems in terms of trigonometric series*, Proceedings of the London Mathematical Society, vol. 34 (1902), pp. 206-221. Cf. also G. D. Birkhoff, *Dynamical systems*, New York (1927), chap. 3, and E. T. Whittaker, *A treatise on the analytical dynamics of particles and rigid bodies*, 4th edition, Cambridge (1937), chap. 16.

[2] Cf. H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Paris (1893), vol. 2.

coefficients of the transformation, it seemed to be probable[3] that the series will diverge in general, but no single example had hitherto been found. From Poincaré's well-known theorem[4] on the analytic integrals of canonical differential equations we can only infer that those series do not uniformly converge, if $\lambda_1, \cdots, \lambda_n$ are variable complex parameters, whereas this theorem cannot be applied to a fixed function H.

We arrange the coefficients of H in a certain order and denote them by $h_1, h_2, h_3, \cdots$. We assume that the power series H converges in a neighbourhood of the origin and that the characteristic roots $\lambda_1, \cdots, \lambda_n$ are pure imaginary linearly independent numbers. The corresponding systems $(h_1, h_2, h_3, \cdots)$ form the points of a space Σ. A point of Σ is called *singular*, if the transformation of H into the normal form (7) *cannot* be performed by a *convergent* contact transformation (4), and else *regular*.

THEOREM 1. *Let* $(c_1, c_2, c_3, \cdots)$ *be a point of* Σ *and* $\epsilon_1, \epsilon_2, \epsilon_3, \cdots$ *an arbitrary sequence of positive numbers. Then a singular point* $(h_1, h_2, h_3, \cdots)$ *of* Σ *exists in the domain*

$$c_k - \epsilon_k < h_k < c_k + \epsilon_k \qquad (k = 1, 2, 3, \cdots).$$

This theorem asserts that the singular points are *everywhere dense* in Σ. We shall reduce the proof to that of another theorem concerning the *integrals* of a canonical system. It would be important to obtain also some information about the distribution of the regular points of Σ, but this seems to be rather a difficult problem. We do not know e.g., if the regular points are also everywhere dense in Σ and if they constitute an open connected set of points. In particular, it would be interesting to decide, whether H is regular or singular in the special case of the restricted problem of three bodies, with respect to the equilibrium solutions of Lagrange. But this seems to be beyond the power of the known methods of analysis.

2. Integrals

If P is any convergent or divergent power series of the $2n$ variables $x_1, \cdots, y_n$, we define the Poisson bracket (P, H) by the power series

$$(P, H) = \sum_{k=1}^{n} (P_{x_k} H_{y_k} - P_{y_k} H_{x_k}).$$

[3] G. D. Birkhoff, *Surface transformations and their dynamical applications*, Acta Mathematica, vol. 43 (1922), pp. 1–119. Cf. on the other hand G. W. Hill, *Remarks on the progress of celestial mechanics since the middle of the century*, Bulletin of the American Mathematical Society, 2nd series, vol. 2 (1896), pp. 125–136, and E. T. Whittaker, *On the adelphic integral of the differential equations of dynamics*, Proceedings of the Royal Society of Edinburgh, vol. 37 (1918), pp. 95–116.

[4] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Paris (1892), vol. 1, chap. 5.

We call P an *integral* of the canonical system (1), if the equation

$$(P, H) = 0$$

holds identically in the variables; in other words, P is an integral, if the relationship

$$\dot{P} = 0$$

follows from (1).

The expression (P, H) is invariant under any contact transformation. By (6) and (8), we obtain

$$(\zeta_k, H) = 0 \qquad (k = 1, \cdots, n).$$

Introducing into $\zeta_k = \xi_k\eta_k$ the original variables $x_1, \cdots, y_n$, we find n power series

$$\zeta_k = P_k(x_1, \cdots, y_n) \qquad (k = 1, \cdots, n) \tag{9}$$

which are integrals of (1). Since the functional determinant of $\xi_1, \cdots, \eta_n$ as functions of $x_1, \cdots, y_n$ does not vanish identically, these n power series are certainly independent one from another, i.e. there exists no power series of the variables P_k with constant coefficients not all zero, which vanishes identically in the variables $x_1, \cdots, y_n$.

Obviously the function H itself and more generally any convergent power series in the single variable H is a *convergent* integral. If H is regular, in the sense of our former definition, there will exist a convergent contact transformation reducing H to the normal form (7), hence the integrals (9) will then also converge. Moreover, by (4), we deduce easily that the integral P_k cannot be expressed as a power series in H alone. Therefore Theorem 1 is contained in

THEOREM 2. *Let $(c_1, c_2, c_3, \cdots)$ be a point of Σ and $\epsilon_1, \epsilon_2, \epsilon_3, \cdots$ an arbitrary sequence of positive numbers. There exists a point $(h_1, h_2, h_3, \cdots)$ in the domain*

$$c_k - \epsilon_k < h_k < c_k + \epsilon_k \qquad (k = 1, 2, 3, \cdots)$$

such that any convergent integral of the corresponding canonical system (1) *is a power series in the single variable H.*

The proof of this theorem depends upon several lemmata.

LEMMA 1. *Any integral of* (1) *can be represented as a power series in the n integrals $P_1, \cdots, P_n$.*

PROOF: It is obvious that the sum, the difference, the product of two integrals and more generally any power series in a finite number of integrals without constant terms is again an integral. Let $P(x_1, \cdots, y_n)$ be any integral of (1). By the contact transformation (4), this integral becomes a power series in the variables $\xi_1, \cdots, \eta_n$ with the general term

$$c_{\alpha_1 \cdots \beta_n} \prod_{k=1}^{n} (\xi_k^{\alpha_k} \eta_k^{\beta_k}),$$

where $c_{\alpha_1 \cdots \beta_n}$ denotes a constant. Consider now the n differences

$$\alpha_k - \beta_k = g_k \qquad (k = 1, \cdots, n).$$

The sum of all special terms with $g_1 = 0, \cdots, g_n = 0$ is a power series T in the n products $\xi_k \eta_k = \zeta_k = P_k(x_1, \cdots, y_n)$, and the expression

$$P - T = J$$

is again an integral. If J is not identically zero, take a term j of least degree $(\alpha_1 + \beta_1) + \cdots + (\alpha_n + \beta_n) = d$ and calculate the terms of degree d in the power series (J, H). Since

$$(J, H) = 0,$$

we obtain from (6) and (7) the relationship

$$\sum_{k=1}^{n} \lambda_k (\xi_k j_{\xi_k} - \eta_k j_{\eta_k}) = 0$$

or

$$\sum_{k=1}^{n} \lambda_k g_k = 0.$$

This is impossible, the numbers $\lambda_1, \cdots, \lambda_n$ being linearly independent and the integers g_k not all zero. Hence $J = 0$ and $P = T$ a power series in the integrals $P_1, \cdots, P_n$.

3. Linear transformation

For our further purposes it is practical to introduce new variables $u_1, \cdots, v_n$ by a special *linear* contact transformation, which reduces the *quadratic* terms of $H(x_1, \cdots, y_n)$ to the normal form $\lambda_1 u_1 v_1 + \cdots + \lambda_n u_n v_n$. Obviously we find such a transformation, if we replace in (4) the power series w by the sum of its quadratic terms. Let

$$z_k = \sum_{l=1}^{n} (c_{kl} u_l + c_{k,l+n} v_l) \qquad (k = 1, \cdots, 2n) \tag{10}$$

be this transformation,

$$\mathfrak{C} = (c_{kl}) \tag{11}$$

its matrix and

$$H = E(u_1, \cdots, v_n) = \sum_{k=1}^{n} \lambda_k u_k v_k + \cdots \tag{12}$$

the power series for H as a function of the new variables. By (3) and (12), the transformed canonical system is

$$\dot{u}_k = E_{v_k} = \lambda_k u_k + \cdots, \quad \dot{v}_k = -E_{u_k} = \lambda_{k+n} v_k + \cdots \quad (k = 1, \cdots, n). \tag{13}$$

Let $\mathfrak{A}$ be the matrix of the coefficients a_{kl} in (2) and $\mathfrak{D}$ the diagonal matrix with the diagonal elements $\lambda_1, \cdots, \lambda_{2n}$. By (2), (10), (11), (13), we deduce

$$\mathfrak{A}\mathfrak{C} = \mathfrak{C}\mathfrak{D}$$

or more explicitly

$$\mathfrak{A}\mathfrak{c}_k = \mathfrak{c}_k\lambda_k \qquad (k = 1, \cdots, 2n),$$

where $\mathfrak{c}_k$ denotes the k^{th} column of the matrix $\mathfrak{C}$. Since $\lambda_1, \cdots, \lambda_n$ are linearly independent, the $2n$ characteristic roots $\lambda_1, \cdots, \lambda_{2n}$ are certainly different one from another. Therefore the general solution of the linear equation

$$\mathfrak{A}\mathfrak{x} = \mathfrak{x}\lambda_k$$

is

$$\mathfrak{x} = \mathfrak{c}_k\rho$$

with an arbitrary scalar factor ρ. On the other hand, λ_k is pure imaginary, and hence by (3), the bar denoting the passage to conjugate complex numbers,

$$\mathfrak{A}\bar{\mathfrak{c}}_k = \bar{\mathfrak{c}}_k\lambda_{k+n} \qquad (k = 1, \cdots, n).$$

This proves the relationship

$$\bar{\mathfrak{c}}_k = \mathfrak{c}_{k+n}\rho_k \qquad (k = 1, \cdots, n); \tag{14}$$

with a certain scalar factor ρ_k. Hence the linear functions z_k of the variables $u_1, \cdots, v_n$ are not changed, if $\mathfrak{c}_l$ is replaced by $\bar{\mathfrak{c}}_l$ $(l = 1, \cdots, 2n)$ and at the same time the variables u_l, v_l by $\hat{u}_l, \hat{v}_l$, where

$$\hat{u}_l = \rho_l^{-1}v_l, \qquad \hat{v}_l = \bar{\rho}_l u_l \qquad (l = 1, \cdots, n); \tag{15}$$

For any power series F, we denote by $\bar{F}$ the power series with the conjugate complex coefficients and the same variables. Since H is a real function,

$$H(x_1, \cdots, y_n) = \bar{H}(x_1, \cdots, y_n),$$

we obtain

$$E(u_1, \cdots, v_n) = \bar{E}(\hat{u}_1, \cdots, \hat{v}_n)$$

and in particular, by (12) and (15),

$$\lambda_k = \bar{\lambda}_k\bar{\rho}_k\rho_k^{-1} \qquad (k = 1, \cdots, n).$$

The characteristic roots $\lambda_1, \cdots, \lambda_n$ are pure imaginary; hence the same holds for the n numbers $\rho_1, \cdots, \rho_n$. Obviously the canonical form of the system (13) and the quadratic terms of E are unchanged, if the variables u_k, v_k are replaced by $a_k^{-1}u_k, a_kv_k$ $(k = 1, \cdots, n)$, where $a_1, \cdots, a_n$ are arbitrary constants $\neq 0$. Denoting the columns $\mathfrak{c}_ka_k^{-1}$, $\mathfrak{c}_{k+n}a_k$ again by $\mathfrak{c}_k$, $\mathfrak{c}_{k+n}$, we have to replace the factor ρ_k by $\rho_k(a_k\bar{a}_k)^{-1}$. Therefore we may assume that $\rho_k = \pm i$.

If $\rho_k = +i$ for any value of k, we replace moreover u_k, v_k by v_k, $-u_k$ and obtain the case $\rho_k = -i$. Hence it is allowed to assume

$$\rho_k = -i \qquad (k = 1, \cdots, n).$$

Now we change the notation. We suppose that the function $H = H^*(x_1, \cdots, y_n)$ and the matrix $\mathfrak{C}$ are given and denote the characteristic roots $\lambda_1, \cdots, \lambda_n$ by $\lambda_1^*, \cdots, \lambda_n^*$, the corresponding power series E by E^*. Let P be the set of all convergent power series

$$E(u_1, \cdots, v_n) = \sum_{k=1}^{n} \lambda_k u_k v_k + \cdots \tag{16}$$

satisfying the condition

$$E(u_1, \cdots, v_n) = \bar{E}(iv_1, \cdots, iu_n),$$

with linearly independent values of $\lambda_1, \cdots, \lambda_n$. Then the inverse of the linear transformation (10) takes any such function E again into a real power series H of the variables $x_k = z_k$, $y_k = z_{k+n}$ $(k = 1, \cdots, n)$.

If F is any power series in the variables $u_1, \cdots, v_n$, we denote by F_l $(l = 0, 1, 2, \cdots)$ the sum of its terms of order l and by $\overline{|F_l|}$ the maximum of the absolute values of the coefficients of these terms. The proof of Theorem 2 is now tantamount to the proof of

THEOREM 3. *Let $\epsilon_2, \epsilon_3, \epsilon_4, \cdots$ be an arbitrary sequence of positive numbers. There exists a power series E of* P *such that*

$$\overline{|E_l - E_l^*|} < \epsilon_l \qquad (l = 2, 3, \cdots)$$

and any convergent integral of the canonical system

$$\dot{u}_k = E_{v_k}, \qquad \dot{v}_k = -E_{u_k} \qquad (k = 1, \cdots, n)$$

is a power series in the single variable E.

Let E be an arbitrary power series of the set P and $H(x_1, \cdots, y_n)$ the corresponding real function. By the linear transformation (10), the n power series

$$\zeta_k = P_k(x_1, \cdots, y_n) \qquad (k = 1, \cdots, n) \tag{17}$$

become integrals of the system

$$\dot{u}_k = E_{v_k}, \qquad \dot{v}_k = -E_{u_k} \qquad (k = 1, \cdots, n). \tag{18}$$

Let Q be one of these integrals and

$$Q_2 = \sum_{k,l=1}^{n} (\alpha_{kl} u_k u_l + \beta_{kl} u_k v_l + \gamma_{kl} v_k v_l)$$

the sum of its terms of second order, where

$$\alpha_{kl} = \alpha_{lk}, \qquad \gamma_{kl} = \gamma_{lk} \qquad (k, l = 1, \cdots, n).$$

Calculating the terms of second order on the left-hand side of the equation

$$(Q, E) = 0,$$

we find by (16) the expression

$$\sum_{k,l=1}^{n} \{\alpha_{kl}(\lambda_k + \lambda_l)u_k u_l + \beta_{kl}(\lambda_k - \lambda_l)u_k v_l - \gamma_{kl}(\lambda_k + \lambda_l)v_k v_l\}.$$

Hence

$$\alpha_{kl} = 0, \quad \gamma_{kl} = 0 \qquad (k, l = 1, \cdots, n),$$

$$\beta_{kl} = 0 \qquad (k \neq l; k, l = 1, \cdots, n),$$

and ζ_k takes the special form

$$\zeta_k = \sum_{l=1}^{n} b_{kl} u_l v_l + \cdots \qquad (k = 1, \cdots, n). \tag{19}$$

By (4) and (10), the variables $\xi_1, \cdots, \eta_n$ can be expressed as power series in $u_1, \cdots, v_n$. The determinant of the linear terms in these power series is different from zero. Let $\psi_1, \cdots, \psi_n$ be indeterminates and

$$\chi_l = \sum_{k=1}^{n} b_{kl}\psi_k \qquad (l = 1, \cdots, n). \tag{20}$$

By (6), (19), (20), the quadratic form

$$\sum_{k=1}^{n} \psi_k \xi_k \eta_k$$

of the $2n$ variables $\xi_1, \cdots, \eta_n$ has the same rank as the quadratic form

$$\sum_{l=1}^{n} \chi_l u_l v_l$$

of the $2n$ variables $u_1, \cdots, v_n$. Hence the n linear forms (20) vanish simultaneously if and only if $\psi_k = 0$ $(k = 1, \cdots, n)$, and consequently their determinant is different from zero. On solving for $u_l v_l$ from (19) we obtain n integrals of (18) in the form

$$S^{(l)} = u_l v_l + \cdots \qquad (l = 1, \cdots, n).$$

Let us suppose that $S^{(l)}$ contains a term of the special type

$$c \prod_{k=1}^{n} (u_k v_k)^{\alpha_k}, \qquad \alpha_1 + \cdots + \alpha_n = g > 1 \tag{21}$$

and that the degree g of this term is as small as possible. Then the integral

$$S^{(l)} - c \prod_{k=1}^{n} S^{(k)\,\alpha_k}$$

does not contain a term of this special type of degree $\leqq g$. It is now obvious that we can also construct n integrals

$$s^{(l)} = u_l v_l + \cdots \qquad (l = 1, \cdots, n),$$

which do not contain *any* term of the form (21). Moreover the integrals $S^{(1)}, \cdots, S^{(n)}$ and consequently the integrals (17) are power series in $s^{(1)}, \cdots, s^{(n)}$. Lemma 1 gives the result, that any integral of (18) is a power series of the variables $s^{(1)}, \cdots, s^{(n)}$. This holds in particular for the integral E. By (16), the power series expressing E as a function of $s^{(1)}, \cdots, s^{(n)}$ has the form

$$E = \sum_{k=1}^{n} \lambda_k s^{(k)} + \cdots;$$

hence we may also represent $s^{(n)}$ as a power series in $E, s^{(1)}, \cdots, s^{(n-1)}$. Therefore we have

LEMMA 2. *There exist $n - 1$ integrals*

$$s^{(l)} = u_l v_l + \cdots \qquad (l = 1, \cdots, n-1)$$

containing no term of the type (21) *and such that any integral of* (18) *is a power series in $E, s^{(1)}, \cdots, s^{(n-1)}$.*

For the rest of our investigation, we will only consider the case of two degrees of freedom. As a matter of fact, the generalization of our proof to the case $n > 2$ requires more complicated calculations, but does not present more serious difficulties.

4. Estimation of the coefficients

LEMMA 3. *Let ξ, η be complex numbers and G a homogeneous polynomial of degree r in the variables u_1, u_2, v_1, v_2. Then*

$$(|\xi| + |\eta|)\,\overline{|G|} \leqq (2r+2)\,\overline{|(\xi u_1 v_1 + \eta u_2 v_2)G|}. \tag{22}$$

PROOF: We can write

$$G = \sum_{l=0}^{r} G^{(l)},$$

where $G^{(l)}$ is also homogeneous in the variables u_1, u_2, of degree l. Since $\overline{|G|}$ is the maximum of the values $\overline{|G^{(l)}|}$ $(l = 0, \cdots, r)$ and $\overline{|(\xi u_1 v_1 + \eta u_2 v_2)G|}$ the maximum of the values $\overline{|(\xi u_1 v_1 + \eta u_2 v_2)G^{(l)}|}$, the inequality (22) will certainly hold, if it is true for $G^{(l)}$ instead of G and $l = 0, \cdots, r$. Hence we may suppose

$$G = \sum_{k=0}^{l} \varphi_k u_1^k u_2^{l-k},$$

where φ_k denotes a homogeneous polynomial in v_1, v_2, of degree $r - l$. If $|\xi| > |\eta|$, we interchange ξ, u_1, v_1 and η, u_2, v_2. Consequently, we have only to consider the case

$$|\xi| \leqq |\eta|. \tag{23}$$

Putting

$$\varphi_{-1} = 0, \qquad \varphi_{l+1} = 0,$$

$$\Phi_k = \xi v_1 \varphi_{k-1} + \eta v_2 \varphi_k \qquad (k = 0, \cdots, l+1),$$

we obtain

$$(\xi u_1 v_1 + \eta u_2 v_2) G = \sum_{k=0}^{l+1} \Phi_k u_1^k u_2^{l+1-k}$$

$$(\eta v_2)^{k+1} \varphi_k = \sum_{p=0}^{k} (-\xi v_1)^{k-p} (\eta v_2)^p \Phi_p \qquad (k = 0, \cdots, l),$$

whence, by (23),

$$|\eta| \overline{|\varphi_k|} \leqq \sum_{p=0}^{k} \overline{|\Phi_p|}.$$

Since $\overline{|G|}$ is now the maximum of the numbers $\overline{|\varphi_k|}$ $(k = 0, \cdots, l)$ and $\overline{|(\xi u_1 v_1 + \eta u_2 v_2) G|}$ the maximum of the numbers $\overline{|\Phi_k|}$ $(k = 0, \cdots, l+1)$, the inequality

$$|\eta| \overline{|G|} \leqq (l+1) \overline{|(\xi u_1 v_1 + \eta u_2 v_2) G|}$$

holds. Moreover

$$|\xi| + |\eta| \leqq 2|\eta|, \qquad l + 1 \leqq r + 1,$$

and the lemma is proved.

By Lemma 2, the existence of $n-1$ integrals $s^{(l)}$ $(l = 1, \cdots, n-1)$ with certain properties was stated. In our case $n = 2$, let us denote the integral $s^{(1)}$ more shortly by s. Then

$$s = \sum_{k=2}^{\infty} s_k, \qquad s_2 = u_1 v_1, \tag{24}$$

where s_k is a homogeneous polynomial in u_1, u_2, v_1, v_2, of degree k. If $k > 2$, the polynomial s_k does not contain a term of the special form $c(u_1 v_1)^\alpha (u_2 v_2)^\beta$.

LEMMA 4. *Let the canonical system*

$$\dot{u}_k = E_{v_k}, \qquad \dot{v}_k = -E_{u_k} \qquad (k = 1, 2) \tag{25}$$

possess a convergent integral, which is not a power series in E alone. Then the sequence

$$\frac{\log \overline{|s_k|}}{k \log k} \qquad (k = 2, 3, \cdots)$$

has a finite upper bound.

PROOF: By Lemma 2, any integral $P(u_1, u_2, v_1, v_2)$ of (25) can be written as a power series in E and s,

$$P = \sum_{\alpha,\beta=0}^{\infty} c_{\alpha\beta} s^\alpha E^\beta.$$

We assume that there exists at least one coefficient

$$c_{\alpha\beta} \neq 0, \qquad \alpha > 0;$$

take $\alpha + \beta = g$ as small as possible. If P is a *convergent* integral, the same holds for the expression

$$p(u_1, u_2, v_1, v_2) = P - \sum_{\beta=0}^{g-1} c_{0\beta} E^{\beta}.$$

By (16) and (24), we find the decomposition

$$p = \sum_{k=2g}^{\infty} p_k,$$

where

$$p_{2g} = \sum_{\alpha+\beta=g} c_{\alpha\beta} s_2^{\alpha} E_2^{\beta}, \qquad s_2 = u_1 v_1, \qquad E_2 = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2.$$

The polynomial

$$\Delta = \frac{\partial p_{2g}}{\partial s_2} = \sum_{\alpha+\beta=g} \alpha c_{\alpha\beta} s_2^{\alpha-1} E_2^{\beta} \tag{26}$$

does not vanish identically; since it is homogeneous in the two variables $u_1 v_1$ and $u_2 v_2$, it may be written in the form

$$\Delta = c \prod_{h=1}^{g-1} (\xi^{(h)} u_1 v_1 + \eta^{(h)} u_2 v_2) \tag{27}$$

with constants $\xi^{(h)}$, $\eta^{(h)}$ not both zero ($h = 1, \cdots, g-1$) and a constant $c \neq 0$.

We denote by x, y, z any three of the variables u_1, u_2, v_1, v_2. Since p is a power series in s and E, the functional determinant

$$\frac{d(E, p, s)}{d(x, y, z)} = 0,$$

identically in x, y, z. Calculating the terms of degree $h - 3$, we obtain

$$\sum_{\alpha+\beta+\gamma=h} \frac{d(E_\alpha, p_\beta, s_\gamma)}{d(x, y, z)} = 0 \qquad (h \geqq 2g + 4).$$

We apply this relationship for $x, y, z = u_1, u_2, v_1$ and for $x, y, z = u_1, u_2, v_2$. Denoting the corresponding expressions

$$\sum_{\alpha+\beta=l} \frac{d(E_\alpha, p_\beta)}{d(y, z)}, \qquad \sum_{\alpha+\beta=l} \frac{d(E_\alpha, p_\beta)}{d(z, x)}, \qquad \sum_{\alpha+\beta=l} \frac{d(E_\alpha, p_\beta)}{d(x, y)} \qquad (l \geqq 2g + 2)$$

by A_{1l}, A_{2l}, A_{3l} and B_{1l}, B_{2l}, B_{3l}, we find

$$\sum_{l+\gamma=h} \left(A_{1l} \frac{\partial s_\gamma}{\partial u_1} + A_{2l} \frac{\partial s_\gamma}{\partial u_2} + A_{3l} \frac{\partial s_\gamma}{\partial v_1} \right) = 0, \tag{28}$$

$$\sum_{l+\gamma=h} \left(B_{1l} \frac{\partial s_\gamma}{\partial u_1} + B_{2l} \frac{\partial s_\gamma}{\partial u_2} + B_{3l} \frac{\partial s_\gamma}{\partial v_2} \right) = 0. \tag{29}$$

Let μ_1, μ_2, μ_3, μ_4 be certain appropriate positive constants, which do not depend upon the subscript k appearing in the formulas. Since the power series E and p converge, the inequalities

$$\overline{|E_k|} < \mu_1^k \qquad (k \geqq 2),$$

$$\overline{|p_k|} < \mu_1^k \qquad (k \geqq 2g)$$

hold, and consequently

$$\overline{\left|\frac{\partial E_k}{\partial x}\right|} < k\mu_1^k, \qquad \overline{\left|\frac{\partial p_k}{\partial y}\right|} < k\mu_1^k.$$

The polynomial $\dfrac{\partial E_k}{\partial x}$ is a sum of $\binom{k+2}{3}$ terms, hence

$$\overline{\left|\frac{d(E_\alpha, p_\beta)}{d(x, y)}\right|} < 2\alpha\beta\binom{\alpha+2}{3}\mu_1^{\alpha+\beta}.$$

If ψ_k denotes any one of the six polynomials $A_{1k}, \cdots, B_{3k}$, we have

$$\overline{|\psi_k|} < \mu_2^k \qquad (k \geqq 2g+2),$$

whence

$$\overline{\left|\psi_l \frac{\partial s_\gamma}{\partial x}\right|} \leqq \gamma\binom{\gamma+2}{3}\mu_2^l\,\overline{|s_\gamma|}. \tag{30}$$

By (26), the identity

$$\frac{d(E_2, p_{2g})}{d(x, y)} = \Delta\,\frac{d(E_2, s_2)}{d(x, y)}$$

holds and therefore we obtain in the case $l = 2g+2$ for $A_{1l}, \cdots, B_{3l}$ the values $\lambda_2 u_1 v_2 \Delta$, 0, $-\lambda_2 v_1 v_2 \Delta$, 0, $\lambda_2 v_1 u_2 \Delta$, $-\lambda_2 v_1 v_2 \Delta$. From (28), (29), (30), we deduce now the inequality

$$\overline{\left|\lambda_2 v_2 \Delta\left(u_1\frac{\partial s_k}{\partial u_1} - v_1\frac{\partial s_k}{\partial v_1}\right)\right|} + \overline{\left|\lambda_2 v_1 \Delta\left(u_2\frac{\partial s_k}{\partial u_2} - v_2\frac{\partial s_k}{\partial v_2}\right)\right|}$$

$$< \sum_{\gamma=2}^{k-1} (\gamma+3)^4\,\overline{|s_\gamma|}\,\mu_2^{k+2g+2-\gamma}$$

and, by (27) and Lemma 3,

$$\overline{\left|u_1\frac{\partial s_k}{\partial u_1} - v_1\frac{\partial s_k}{\partial v_1}\right|} + \overline{\left|u_2\frac{\partial s_k}{\partial u_2} - v_2\frac{\partial s_k}{\partial v_2}\right|} < k^{g+3}\sum_{\gamma=2}^{k-1}\overline{|s_\gamma|}\,\mu_3^{k-\gamma} \qquad (k \geqq 3). \tag{31}$$

If

$$\omega = a u_1^{\alpha_1} v_1^{\beta_1} u_2^{\alpha_2} v_2^{\beta_2}$$

is any term of s_k, then

$$|\alpha_1 - \beta_1| + |\alpha_2 - \beta_2| \geqq 1,$$

$$u_1 \frac{\partial \omega}{\partial u_1} - v_1 \frac{\partial \omega}{\partial v_1} = (\alpha_1 - \beta_1)\omega, \qquad u_2 \frac{\partial \omega}{\partial u_2} - v_2 \frac{\partial \omega}{\partial v_2} = (\alpha_2 - \beta_2)\omega,$$

and (31) implies

$$\overline{|s_k|} < k^{g+3} \sum_{\gamma=2}^{k-1} \overline{|s_\gamma|}\, \mu_3^{k-\gamma} \qquad (k = 3, 4, \cdots). \tag{32}$$

Obviously the inequality

$$\overline{|s_l|} \leqq (2l^{g+3}\mu_3)^{l-2} \tag{33}$$

is true for $l = 2$. If it is proved for $l = 2, 3, \cdots, k-1$, we find from (32) the relationship

$$\overline{|s_k|} < k^{g+3} \sum_{\gamma=2}^{k-1} (2\gamma^{g+3}\mu_3)^{\gamma-2} \mu_3^{k-\gamma} \leqq (k^{g+3}\mu_3)^{k-2} \sum_{\gamma=2}^{k-1} 2^{\gamma-2} < (2k^{g+3}\mu_3)^{k-2},$$

and (33) holds for $l = k$. Hence

$$\overline{|s_k|} < k^{\mu_4 k}$$

$$\frac{\log \overline{|s_k|}}{k \log k} < \mu_4 .$$

5. Proof of Theorem 3

On account of Lemma 4, it is sufficient for the proof of Theorem 3 to construct a power series

$$E = \sum_{k=2}^{\infty} E_k$$

with the following 4 properties:

I) $$E_k(u_1, u_2, v_1, v_2) = \bar{E}_k(iv_1, iv_2, iu_1, iu_2) \quad (k = 2, 3, \cdots);$$

II) $$E_2 = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2,$$

where λ_1, λ_2 *are linearly independent;*

III) $$\overline{|E_k - E_k^*|} < \epsilon_k \qquad (k = 2, 3, \cdots);$$

IV) *the sequence*

$$\frac{\log \overline{|s_k|}}{k \log k} \qquad (k = 2, 3, \cdots)$$

has no finite upper bound.

The positive numbers $\epsilon_2, \epsilon_3, \cdots$ are arbitrarily given. Obviously we may suppose that

$$\epsilon_k < 1 \qquad (k = 2, 3, \cdots).$$

We begin with the construction of λ_1, λ_2. The coefficients λ_1^*, λ_2^* in

$$E_2^* = \lambda_1^* u_1 v_1 + \lambda_2^* u_2 v_2$$

are pure imaginary and linearly independent; hence

$$\omega^* = \frac{\lambda_1^*}{\lambda_2^*}$$

is a real number. We choose two integers q, r such that

$$q > 1 + |\omega^*| + 2|\lambda_2^*|\epsilon_2^{-1}, \tag{34}$$

$$|q\omega^* + r| < 1 \tag{35}$$

and define three sequences of numbers q_m, r_m, l_m $(m = 1, 2, \cdots)$ in the following manner:

$$r_m = q_m\left(\frac{r}{q} - \sum_{k=1}^{m} q_k^{-1}\right); \tag{36}$$

$$l_m = q_m + |r_m|; \tag{37}$$

$$q_1 = q^2; \tag{38}$$

q_{m+1} is the least integral power of q satisfying the inequality

$$q_{m+1} > q_m^2 + 4|\lambda_2^*|\epsilon_{l_m}^{-1} l_m^{l_m}. \tag{39}$$

It is obvious that the numbers q_m, r_m, l_m are uniquely determined and q_m, l_m are positive. For the exponent a_m in

$$q_m = q^{a_m}$$

the inequality

$$a_{m+1} > 2a_m$$

holds. Since $a_1 = 2$, we obtain

$$a_m \geqq 2^m, \qquad a_{m+1} - a_m > a_m \geqq 2^m, \tag{40}$$

hence r_m and l_m are integers and the sequence $q_m q_{m+1}^{-1}$ tends to zero. By (34), (38), (40), the series

$$\theta = \sum_{k=1}^{\infty} q_k^{-1} \tag{41}$$

converges, and we find the estimations

$$0 < \sum_{k=h}^{\infty} q_k^{-1} < q_h^{-1} \sum_{k=0}^{\infty} q^{-k} \leqq 2q_h^{-1} \qquad (h = 1, 2, \dots) \tag{42}$$

$$0 < \theta < 2q_1^{-1} \leqq q^{-1}. \tag{43}$$

Moreover, by (35), (36), (41),

$$|r_m| < q_m(|\omega^*| + q^{-1} + \theta),$$

whence, by (34), (37), (38), (39), (43),

$$l_{m+1} - l_m \geqq q_{m+1} - q_m - |r_m| > q_m(q^2 - 1 - |\omega^*| - 2q^{-1}) > 0. \tag{44}$$

Let λ_1, λ_2 be defined by

$$\lambda_1 = \lambda_2^*\left(\theta - \frac{r}{q}\right), \qquad \lambda_2 = \lambda_2^*; \tag{45}$$

then λ_1, λ_2 are pure imaginary and

$$E_2 = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2$$

satisfies I) with $k = 2$. For the expression

$$\rho_m = \lambda_1 q_m + \lambda_2 r_m \qquad (m = 1, 2, \dots) \tag{46}$$

we find, by (36), (41), (45),

$$\rho_m = \lambda_2 q_m \sum_{k=m+1}^{\infty} q_k^{-1}$$

and by (42)

$$0 < \rho_m \lambda_2^{-1} < 2q_m q_{m+1}^{-1}. \tag{47}$$

If x, y are two indeterminates, the identity

$$(\lambda_1 x + \lambda_2 y)q_m\lambda_2^{-1} = (q_m y - r_m x) + \rho_m \lambda_2^{-1} x$$

holds. For given integral values of x and y, the number $q_m y - r_m x$ is integral, whereas the absolute value of $\rho_m\lambda_2^{-1}x$, by (47), is less than 1, if m is sufficiently large, and 0 only in the case $x = 0$. Therefore the numbers λ_1, λ_2 are linearly independent. On the other hand, by (34), (35), (43), (45),

$$|\lambda_1 - \lambda_1^*| = |\lambda_2^*|\left|\theta - \frac{r}{q} - \omega^*\right| < |\lambda_2^*|(q^{-1} + q^{-1}) < \epsilon_2.$$

Consequently II) is satisfied, and III) for $k = 2$.

By (34), (37), (38), we have $l_1 \geqq 4$. If we define

$$E_k = E_k^* \qquad (2 < k < l_1),$$

the conditions I) and III) are satisfied for $k < l_1$. Let m be any positive integer and assume, that the polynomials E_k are already determined for $k < l_m$ such that I) and III) hold. Consider now any power series $E = E_2 + E_3 + \cdots$ of P with these fixed terms E_k ($k = 2, 3, \cdots, l_m - 1$). Let s be the integral

$$s = \sum_{k=2}^{\infty} s_k, \qquad s_2 = u_1 v_1$$

of the corresponding canonical system. Then

$$(s, E) = 0,$$

whence

$$\sum_{h=2}^{l} (s_h, E_{l+2-h}) = 0 \qquad (l = 3, 4, \cdots)$$

$$\lambda_1\left(u_1 \frac{\partial s_l}{\partial u_1} - v_1 \frac{\partial s_l}{\partial v_1}\right) + \lambda_2\left(u_2 \frac{\partial s_l}{\partial u_2} - v_2 \frac{\partial s_l}{\partial v_2}\right) = u_1 \frac{\partial E_l}{\partial u_1} - v_1 \frac{\partial E_l}{\partial v_1} - \sum_{h=3}^{l-1} (s_h, E_{l+2-h}). \tag{48}$$

Let $s_{\alpha\beta\gamma\delta} u_1^\alpha u_2^\beta v_1^\gamma v_2^\delta$, $E_{\alpha\beta\gamma\delta} u_1^\alpha u_2^\beta v_1^\gamma v_2^\delta$ be corresponding terms of s_l, E_l, where $s_{\alpha\beta\gamma\delta}$, $E_{\alpha\beta\gamma\delta}$ are the coefficients. Then $\alpha + \beta + \gamma + \delta = l$ and $s_{\alpha\beta\gamma\delta} = 0$ for $\alpha = \gamma$, $\beta = \delta$. From (48) we obtain

$$\{(\alpha - \gamma)\lambda_1 + (\beta - \delta)\lambda_2\} s_{\alpha\beta\gamma\delta} = (\alpha - \gamma) E_{\alpha\beta\gamma\delta} + b_{\alpha\beta\gamma\delta}, \tag{49}$$

where $b_{\alpha\beta\gamma\delta}$ denotes a certain bilinear function of the coefficients of $s_3, \cdots, s_{l-1}$ and $E_3, \cdots, E_{l-1}$. Since E_k is fixed for $k = 3, 4, \cdots, l_m - 1$, we infer from (49), that the coefficients of s_k for $k = 3, 4, \cdots, l_m - 1$ are uniquely determined and that the same holds for the expression

$$s_{\alpha\beta\gamma\delta} - \frac{\alpha - \gamma}{(\alpha - \gamma)\lambda_1 + (\beta - \delta)\lambda_2} E_{\alpha\beta\gamma\delta} \tag{50}$$

with $\alpha + \beta + \gamma + \delta = l_m$ and $\alpha - \gamma$, $\beta - \delta$ not both zero. Take in particular

$$\alpha = q_m, \quad \beta = r_m \quad \gamma = 0, \quad \delta = 0, \quad \text{if} \quad r_m \geqq 0,$$

$$\alpha = q_m, \quad \beta = 0, \quad \gamma = 0, \quad \delta = -r_m, \quad \text{if} \quad r_m < 0$$

and denote the corresponding coefficients of s_{l_m}, E_{l_m}, $E^*_{l_m}$ more shortly by σ_m, η_m, η^*_m. Then, by (46) and (50), the value

$$\sigma_m - q_m \rho_m^{-1} (\eta_m - \eta^*_m) = \nu_m \tag{51}$$

is uniquely given.

If we choose for η_m the two values

$$\eta_m = \eta^*_m \pm \tfrac{1}{2} \epsilon_{l_m}, \tag{52}$$

we find, by (51), two values of σ_m which have the difference $q_m\rho_m^{-1}\epsilon_{l_m}$. By (39), (45), (47), the inequality

$$|\, q_m\rho_m^{-1}\epsilon_{l_m}\,| > \tfrac{1}{2}q_{m+1}\,|\,\lambda_2\,|^{-1}\epsilon_{l_m} > 2l_m^{ml_m}$$

holds. Hence we can determine the sign in (52) such that

$$|\,\sigma_m\,| > l_m^{ml_m}. \tag{53}$$

By (44), we have $l_m < l_{m+1}$. For $k = l_m, l_m + 1, \cdots, l_{m+1} - 1$ we define E_k by

$$E_{l_m} = E_{l_m}^* \pm \tfrac{1}{2}\epsilon_{l_m}(u_1^{q_m}u_2^{r_m} + i^{l_m}v_1^{q_m}v_2^{r_m}), \qquad \text{if } r_m \geqq 0,$$

$$E_{l_m} = E_{l_m}^* \pm \tfrac{1}{2}\epsilon_{l_m}(u_1^{q_m}v_2^{-r_m} + i^{l_m}v_1^{q_m}u_2^{-r_m}), \qquad \text{if } r_m < 0,$$

$$E_k = E_k^* \qquad (l_m < k < l_{m+1}).$$

Now the conditions I) and III) are also satisfied for $l_m \leqq k < l_{m+1}$.

By this construction, a power series E satisfying I), II), III) is uniquely determined. The corresponding integral

$$s = \sum_{k=2}^{\infty} s_k$$

contains in the term s_{l_m} the coefficient σ_m. By (53), the inequality

$$|\,s_{l_m}\,| > l_m^{ml_m} \qquad (m = 1, 2, \cdots)$$

holds. Hence the condition IV) is also satisfied, and the theorem is proved.

INSTITUTE FOR ADVANCED STUDY.

An extra number of the Annals of Mathematics will be published within a few months. This will be number 5 of volume 42. This extra number has been made possible by the generosity of the American Philosophical Society. The Index for the entire volume 42 will be incorporated in the extra number.